



Brussel, **XXX**  
SWD(2014) **XXX**

**WERKDOCUMENT VAN DE DIENSTEN VAN DE COMMISSIE**

**Gemeenschappelijke methodiek voor de evaluatie van staatssteun**

**WERKDOCUMENT VAN DE DIENSTEN VAN DE COMMISSIE**

**Gemeenschappelijke methodiek voor de evaluatie van staatssteun**

## Inhoudsopgave

|              |  |    |
|--------------|--|----|
| 1            | Inleiding .....  | 4  |
| 2            | De doelstellingen van de evaluatie van staatssteun .....   | 5  |
| 3            | Het evaluatieplan .....  | 7  |
| 3.1          | Te evalueren doelstellingen van de steunregeling.....  | 7  |
| 3.2          | De evaluatievragen.....  | 7  |
| 3.3          | Resultaatindicatoren .....   | 9  |
| 3.4          | Methoden: een geschikte vergelijkingsbasis zoeken .....  | 9  |
| 3.5          | Verzamelen van data: de best mogelijke bronnen gebruiken .....   | 11 |
| 3.6          | Tijdpad evaluatie .....  | 13 |
| 3.7          | De evaluerende instantie: onafhankelijkheid en deskundigheid verzekeren .....                                  | 14 |
| 3.8          | Ruchtbaarheid geven aan de evaluatie: de betrokkenheid van stakeholders stimuleren....                         | 15 |
| 4            | Criteria voor de selectie van te evalueren steunregelingen .....   | 16 |
| 4.1          | Grote steunregelingen, ook steunregelingen in het kader van de algemene<br>groepsvrijstellingsverordening..... | 16 |
| 4.2          | Vernieuwende steunregelingen .....   | 17 |
| 4.3          | Steunregelingen beïnvloed door te verwachten ingrijpende veranderingen .....                                   | 17 |
| 4.4          | Andere steunregelingen .....   | 18 |
| Bijlage I:   | Technische bijlage over relevante methoden om de causale effecten te identificeren ...                         | 19 |
| Bijlage II:  | Lijst van mogelijke resultaatindicatoren .....   | 40 |
| Bijlage III: | Verklarende woordenlijst.....  | 45 |
| Bijlage IV:  | Literatuurlijst .....  | 46 |

## 1 Inleiding

Lidstaten verlenen staatssteun om sterk uiteenlopende beleidsdoelstellingen te helpen bereiken, zoals het afbouwen van regionale verschillen binnen een land, het bevorderen van onderzoek, ontwikkeling en innovatie (O&O&I), of het stimuleren van een hoog niveau aan milieubescherming.

Welke soorten steun met de interne markt verenigbaar zijn, wordt in de EU-staatssteunregels bepaald op basis van een systeem van onderzoek vooraf: steunregelingen<sup>1</sup> worden goedgekeurd op basis van vooraf bepaalde beoordelingscriteria, in de aanname dat, indien aan die beoordelingscriteria wordt voldaan, de positieve effecten van de steun zullen opwegen tegen eventuele negatieve effecten. Doorgaans vindt deze beoordeling van regelingen plaats zonder dat er een afdoende evaluatie plaatsvindt van de daadwerkelijke impact ervan op markten in de tijd.

Tot dusver is bij het toepassen van EU-staatssteunregels betrekkelijk weinig belang gehecht aan ex-post-bewijsmateriaal over wat daadwerkelijk met de overheidsmiddelen is bereikt of wat de impact van staatssteun op de mededinging is. Toch is het voor beleidsmakers zowel in de lidstaten als op EU-niveau van essentieel belang om rekening te houden met de meetbare resultaten van staatssteun die in het verleden is verleend en met de lessen die daaruit kunnen worden getrokken. Dit zal ervoor helpen te zorgen dat uit staatssteun gefinancierde regelingen doeltreffender worden en de markten minder verstoren. Tegelijk zal hiermee ook de doelmatigheid van nieuwe regelingen en mogelijk zelfs ook van toekomstige regels voor steunverlening verbeteren.

Een aantal landen evalueert nu reeds hun subsidieregelingen, ook al gebeurt dit niet altijd op regelmatige basis.<sup>2</sup> Ook voor EU-uitgaven (met inbegrip van financiering uit de EU-structuur- en investeringsfondsen, zoals het Europees Fonds voor Regionale Ontwikkeling (EFRO), Europees Sociaal Fonds (ESF) en het Europees Landbouwfonds voor plattelandsontwikkeling (ELFPO)) vindt er vooraf, tijdens en achteraf evaluatie plaats, aan de hand van de desbetreffende verordeningen en handreikingen van de Commissie.<sup>3</sup> Om bij de evaluaties door de lidstaten duplicering te vermijden, is in de handreiking "Concepts and

---

1 Steunregelingen zijn goed voor het merendeel van alle toegekende steun: volgens de gegevens van het Scorebord Staatssteun voor 2013 vertegenwoordigen steunregelingen 23% van alle steunmaatregelen en 55% van het totale steunvolume, terwijl daarnaast de categorie van regelingen in het kader van een groepsvrijstelling goed is voor 63% van alle steunmaatregelen en rond 32% van het totale steunvolume. Verordening (EG) nr. 659/1999 omschrijft "steunregeling" als "elke regeling op grond waarvan aan ondernemingen die in de regeling op algemene en abstracte wijze zijn omschreven, individuele steun kan worden toegekend zonder dat hiervoor nog uitvoeringsmaatregelen vereist zijn, alsmede elke regeling op grond waarvan steun die niet gebonden is aan een specifiek project voor onbepaalde tijd en/of voor een onbepaald bedrag aan een of meer ondernemingen kan worden toegekend".

2 Zo worden in diverse lidstaten op regelmatige basis evaluatierapporten opgesteld ten behoeve van de nationale rekenkamer of het nationale parlement.

3 De betrokken concepten en aanbevelingen komen in detail aan bod in de handreikingen van de Commissie met betrekking tot evaluatie voor de financieringsperiode 2014-2020 (beschikbaar onder: [http://ec.europa.eu/regional\\_policy/information/evaluations/guidance\\_en.cfm](http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm)).

Recommendations" met betrekking tot monitoring en evaluatie duidelijk aangegeven dat aan de eisen van de Europese structuur- en investeringsfondsen op het gebied van evaluatie kan worden voldaan door de evaluaties uit te voeren die door de staatssteunregels worden verlangd.

Het initiatief voor de modernisering van het EU-staatssteunbeleid (SAM)<sup>4</sup> wil de handhavingsinspanningen van de Commissie concentreren op grotere steunregelingen die waarschijnlijk de aanzienlijkste effecten op de interne markt zullen hebben. Tegelijk zal de analyse van meer lokale zaken met minder of beperktere effecten op het handelsverkeer, worden vereenvoudigd, onder meer door lidstaten meer flexibiliteit te bieden bij de tenuitvoerlegging van dit soort maatregelen. Daartoe wordt het toepassingsgebied van de nieuwe algemene groepsvrijstellingsverordening verruimd<sup>5</sup>. Om ervoor te zorgen dat, per saldo, de positieve effecten van staatssteun (doordat deze het gestelde doel vervult) blijven opwegen tegen de potentieel negatieve effecten op de mededinging en het handelsverkeer, en om buitensporige verstoring van de markt te voorkomen, dient een grotere vereenvoudiging gepaard te gaan met meer transparantie, versterkt toezicht op de naleving van de staatssteunregels op nationaal en op Europees niveau, en een daadwerkelijke evaluatie<sup>6</sup>.

Dit werkdocument beschrijft een gemeenschappelijke methodiek voor de evaluatie (doorlichting) van staatssteunregelingen. Dit document is bedoeld als handreiking voor overheden die evaluaties plannen en uitvoeren.

## **2 De doelstellingen van de evaluatie van staatssteun**

De algemene doelstelling van de evaluatie van staatssteun is de positieve en negatieve effecten van een regeling tegen elkaar af te zetten: een afweging van de beleidsdoelstelling van de steun tegenover de effecten ervan op de mededinging en het interstatelijke handelsverkeer. De evaluatie van staatssteun kan duidelijk maken of en in hoeverre de oorspronkelijke doelstellingen van een steunregeling zijn bereikt (d.w.z. de positieve effecten meten) en kan bepalen wat de impact van de regeling op de markten en de mededinging was (d.w.z. wat de eventuele negatieve effecten waren). Een evaluatie verschilt dus qua opzet van twee ex-post-exercities die de Commissie nu reeds uitvoert voor staatssteunregelingen: monitoring<sup>7</sup> en verslaglegging<sup>8</sup>.

---

4 Mededeling van de Commissie aan het Europees Parlement, de Raad, het Europees Economisch en Sociaal Comité en het Comité van de Regio's "De modernisering van het EU-staatssteunbeleid" (SAM), COM(2012) 209 final van 8.5.2012.

5 Verordening (EU) nr. 651/2014 van de Commissie van 17 juni 2014 waarbij bepaalde categorieën steun op grond van de artikelen 107 en 108 van het Verdrag met de interne markt verenigbaar worden verklaard, PB L 187 van 26.6.2014, blz. 1.

6 Zie ook de conclusies van de Raad van 13 november 2012 over de hervorming van het staatssteuntoezicht.

7 In de monitoringexercitie onderzoekt de Commissie periodiek een steekproef van door lidstaten ten uitvoer gelegde steunmaatregelen op hun rechtmatigheid. Deze exercitie moet ervoor zorgen dat lidstaten besluiten van de Commissie correct uitvoeren en zich houden aan de desbetreffende rechtsvoorschriften (zoals die zijn vastgelegd in de algemene groepsvrijstellingsverordening). De Commissie gaat voor een representatieve steekproef van zaken ook na in hoeverre de vooraf vastgestelde regels en voorwaarden zijn nageleefd.

Bij de evaluatie van staatssteun zou met name het directe stimulerende effect van de steun op de begunstigde moeten kunnen worden gemeten (de vraag dus of de begunstigde door de steun zijn gedrag heeft gewijzigd en hoe groot de impact van de steun wel is geweest). Ook moet deze evaluatie aanwijzingen opleveren voor de algemene positieve en negatieve effecten van de steunregeling voor het behalen van de beoogde beleidsdoelstelling en voor de mededinging en het handelsverkeer. Zij zou ook kunnen onderzoeken in hoeverre het gekozen steuninstrument evenredig en geschikt was.

Op basis van die meting kan de evaluatie dan bevestiging geven voor de vraag of de aannames die ten grondslag liggen aan de goedkeuring vooraf van de steunregeling, nog steeds gelden en kan zij helpen de vormgeving van toekomstige steunregelingen en staatssteunregels te verbeteren. Voorts kan evaluatie als basis dienen om toekomstige overheidsinterventies bij te sturen zodat de doeltreffendheid en doelmatigheid van de steun zodanig wordt verbeterd dat de positieve effecten voldoende vast staan om de door de interventie veroorzaakte verstoring van de markt te kunnen accepteren. Dit soort verbeteringen van toekomstige regelingen zou kunnen uiteenlopen van aanpassingen aan de vormgeving (met onder meer aangepaste selectiecriteria en een uitgebreidere meting van het stimulerende effect) tot ingrijpendere aanpassingen (zoals het bevorderen van het gebruik van een alternatieve vorm van steun, herdefiniëren van doelstellingen of beoogde begunstigden, of het overwegen om voor dezelfde beleidsdoelstellingen met andere opties dan steun te werken).

Het is belangrijk om voor de evaluatie een goed tijdpad te bepalen, dat voldoende tijd laat om genoeg bewijsmateriaal te verzamelen terwijl beleidsmakers toch zo snel mogelijk resultaten ontvangen, zodat potentiële verbeteringen tijdig kunnen worden aangebracht.<sup>9</sup> Evaluaties van staatssteun dienen daarom normaal gesproken te worden beschouwd als doorlopende evaluaties (die plaatsvinden terwijl de steunregeling nog loopt) en niet als evaluaties die louter achteraf plaatsvinden (wanneer de tenuitvoerlegging van de regeling is voltooid). Daarbij dient rekening te worden gehouden met specifieke gevallen waarin de effecten van een interventie pas op langere termijn volledig zichtbaar worden en waarbij de evaluatie alleen de initiële effecten in beeld kan brengen en meten.

De evaluatie van staatssteun zou uiteindelijk een leerervaring moeten zijn voor zowel de Commissie als de lidstaten. Dit kan alleen gebeuren indien de evaluatie aan een aantal minimale kwaliteitsnormen voldoet. Daarom dient de Commissie ervoor te zorgen dat de nodige kwaliteitscontrole op evaluaties plaatsvindt. Met name zal de Commissie in detail onderzoeken wat de algemene betrouwbaarheid van de evaluatie is en zal zij wijzen op

---

8 De primaire doelstelling van de jaarlijkse verslaglegging door lidstaten is de transparantie te vergroten van staatssteun die lidstaten toekennen. Tegelijk is deze exercitie ook een bron van betrouwbare statistische gegevens die voor beleidsvorming en monitoringdoeleinden kan worden gebruikt. De data in de jaarlijkse verslagen bevatten vooral informatie in kwantitatieve termen (om bijvoorbeeld te laten zien voor welke doelstellingen staatssteun werd verleend en hoe hoog het uitgetrokken budget was). De Commissie gebruikt deze verslagen van de lidstaten om het Scorebord Staatssteun samen te stellen.

9 In een aantal staatssteunrichtsnoeren wordt voor te evalueren steunregelingen vier jaar als normale looptijd genoemd.

mogelijke tekortkomingen in twee cruciale fases: het evaluatieplan (evaluatieontwerp) en het eindrapport. Waar nodig kan de Commissie een beroep doen op externe, onafhankelijke deskundigen om haar bij te staan bij de kwaliteitscontrole op de evaluatie.

De Commissie kan voor nationale overheidsdiensten ook opleidingssessies en workshops organiseren over evaluatiemethoden en -technieken. Daarnaast kunnen succesvolle ervaringen en goede praktijken van lidstaten worden uitgewisseld en gebruikt om in de toekomst doeltreffendere steunregelingen te helpen ontwerpen.

De voordelen van het uitvoeren van evaluaties zullen over een paar jaar duidelijk zichtbaar worden, wanneer de eerste evaluatierapporten gereed komen en de bevindingen en aanbevelingen ervan beschikbaar worden gesteld. Die zullen dan kunnen worden gebruikt om de vormgeving van nieuwe steunregelingen en mogelijk ook de staatssteunregels te verbeteren. Op de middellange tot lange termijn zou evaluatie geleidelijk tot meer fundamentele veranderingen in de algemene benadering van staatssteun kunnen leiden.

### **3 Het evaluatieplan**

Het is van essentieel belang dat in een vroeg stadium - parallel met de vormgeving van de staatssteunregeling - een omvattend plan wordt uitgewerkt voor de evaluatie van die regeling. De goedkeuring van het evaluatieplan door de Commissie is van cruciaal belang om een gelijke behandeling te verzekeren. Dit plan moet nadien strikt worden toegepast.

Algemeen wordt immers erkend dat evaluaties des te doeltreffender zijn wanneer zij vooraf goed zijn gepland en voorbereid, met name omdat het op die manier eenvoudiger wordt om de geschikte gegevens te verzamelen. Door reeds in een vroeg stadium te plannen, zullen waarschijnlijk ook minder middelen nodig zijn om de evaluatie uit te voeren, hetgeen uiteindelijk ook de kwaliteit van de evaluatie ten goede zal komen.

Het evaluatieplan dat de lidstaat, overeenkomstig de desbetreffende regels, bij de Commissie moet aanmelden, moet ten minste de volgende minimale elementen bevatten.

#### **3.1 Te evalueren doelstellingen van de steunregeling**

De eerste fase bij de evaluatie van een regeling is dat duidelijk de onderliggende "interventielogica" van de steunregeling wordt aangegeven. Dit betekent dat een beschrijving wordt gegeven van de behoeften en problemen waarop de regeling een antwoord wil bieden, de beoogde begunstigden en doelstellingen, de algemene en specifieke doelstellingen en de verwachte effecten. Ook moeten de belangrijkste aannames met betrekking tot externe factoren worden vermeld die invloed op de regeling kunnen hebben.

#### **3.2 De evaluatievragen**

Het evaluatieplan moet aangeven hoe ruim de evaluatie zal zijn, het moet dus precieze vragen bevatten die kwantitatief te beantwoorden zijn en vergezeld gaan van de nodige documenten

ter staving. Deze evaluatievragen moeten de klemtoon leggen op de effecten van de steunregeling en kunnen op de volgende drie niveaus worden ingedeeld:

1. Directe impact van de steun op de begunstigden ervan - met vragen als:

- Heeft de steun een aanzienlijk effect gehad op de gedragslijn van de begunstigden van de steun? (stimulerend effect)
- Heeft de steun een effect op de situatie van de begunstigden ervan? (Is bijvoorbeeld hun concurrentiepositie of hun wanbetalingsrisico veranderd?)
- In hoeverre heeft de steun de ervan verwachte effecten opgeleverd?
- Waren er verschillen in de effecten van de steun op de begunstigden ervan? (bijvoorbeeld afhankelijk van hun grootte, locatie of sector)

2. Indirecte impact van de steunregeling - met vragen als:

- Heeft de regeling overloopeffecten gehad naar de activiteiten van andere ondernemingen of naar andere geografische regio's? Heeft de steun investeringen van concurrenten weggedrukt of activiteiten aangetrokken uit locaties in de buurt?
- Heeft de regeling bijgedragen tot de desbetreffende beleidsdoelstelling?
- Zijn de geaggregeerde effecten op de mededinging en het handelsverkeer meetbaar?

3. Evenredigheid en geschiktheid van de steunregeling - met vragen als:

- Was de steunregeling evenredig aan het aan te pakken probleem? Konden dezelfde effecten worden bereikt met minder steun of een andere vorm van steun? (bijv. leningen in plaats van subsidies)
- Werd voor het meest doeltreffende steuninstrument gekozen? Zouden andere steuninstrumenten of andere soorten interventies beter geschikt zijn om de betrokken doelstelling te behalen?

De evaluatie dient, voor zover mogelijk, de impact van de steunregeling op alle drie de niveaus te meten, waarbij wordt gekeken naar de voor de doelstellingen van de regeling relevante vragen. Toch is de directe impact van steun op de begunstigden doorgaans het soort impact dat robuust meetbaar is. In de praktijk zijn de meeste evaluatiemethoden ontwikkeld om dit soort impact te meten. Voorts is de evaluatie van de directe effecten van de steun (zoals het stimulerende effect ervan) van kapitaal belang, omdat deze een waardevol inzicht bieden in de soorten indirecte effecten en verstoringen die te verwachten vallen. Meer bepaald kan, wanneer de steun geen stimulerend effect heeft, worden aangenomen dat de steun verstorend is, omdat hij de begunstigden ervan *windfall gains* oplevert.



### 3.3 Resultaatindicatoren

De evaluatievragen zouden moeten leiden tot de keuze van specifieke resultaatindicatoren die kwantitatieve gegevens opleveren over de resultaten die met de staatssteunregeling worden geboekt. In bijlage II is een indicatieve en niet-uitputtende lijst van resultaatindicatoren te vinden voor zowel de directe als de indirecte impact van een regeling, met inbegrip van de mogelijke effecten voor de mededinging en het handelsverkeer. De resultaatindicatoren zullen afhankelijk zijn van de doelstellingen van de te evalueren steunmaatregel. Het evaluatieplan moet aangeven waarom de gekozen indicatoren het meest relevant zijn voor het meten van de impact van deze steunregeling.

### 3.4 Methoden: een geschikte vergelijkingsbasis zoeken

Evaluaties van staatssteun zouden de causale effecten van de regeling zelf in beeld moeten brengen, zonder dat dit beeld vertekend wordt door andere variabelen die een effect kunnen hebben gehad op de geobserveerde uitkomst zoals algemene macro-economische omstandigheden of de heterogeniteit van de ondernemingen (bijv. verschillen in grootte, locatie, financiële middelen of managementcapaciteit). In het evaluatieplan dienen de belangrijkste methoden te zijn opgenomen die zullen worden gebruikt om het effect van de steun te bepalen. Ook dient te worden besproken waarom mag worden verwacht dat die methoden geschikt zullen zijn voor de betrokken regeling.

Dit causale effect is het verschil tussen de uitkomst mét steun en de uitkomst zonder steun. Daar waar de uitkomst mét steun is vast te stellen voor ondernemingen die de steun ontvangen, valt de uitkomst zonder steun alleen te meten voor ondernemingen die geen steun ontvangen. Per definitie valt niet te observeren wat de uitkomst zonder steun was geweest voor de ondernemingen die wel steun hebben ontvangen. Om het effect van de steun op begunstigden van de steun te ramen, is het dus nodig om dit nulscenario (*counterfactual*) te construeren, op basis van de best vergelijkbare onderneming(en) of een controlegroep.

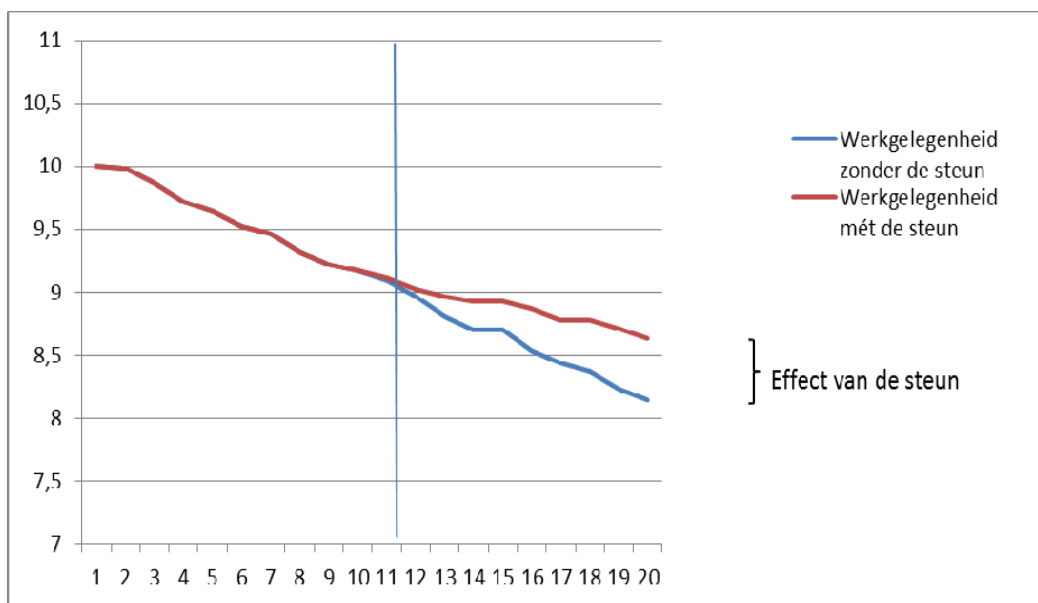
De kwaliteit van deze controlegroep is van cruciaal belang voor de validiteit van de evaluatie.

Ondernemingen die steun krijgen, kunnen zeer wel in een andere situatie verkeren dan ondernemingen die geen steun krijgen. Zij hebben misschien te maken met andere lokale vraag- en aanbodomstandigheden, hebben minder gemakkelijk toegang tot krediet, of zijn meer of minder doelmatig. Al deze factoren kunnen een effect hebben op het prestatie- of activiteitspeil van de ondernemingen, zowel wanneer zij steun ontvangen als wanneer dat niet het geval is. De prestaties van begunstigden vergelijken met die van niet-begunstigden zal waarschijnlijk meer déze realiteit in beeld brengen dan het effect van de steun zelf. Een evaluatie van de steunregeling mag dus niet berusten op een eenvoudige vergelijking van begunstigden met niet-begunstigden, maar moet rekening houden met de verschillende kenmerken van de beide groepen ondernemingen, zowel die welke observeerbaar zijn als die waarvoor dat niet geldt.

In het geval van bijvoorbeeld regionale steun presteren begunstigden van de steun in regio's waar de marktomstandigheden ongunstig zijn (d.w.z. waar de lokale product-, arbeids- of

kapitaalmarkten zwak presteren), doorgaans slechter dan niet-begunstigden in meer welvarende regio's. Een en ander brengt echter in geen geval het effect van de steun tot uiting. De relevante vraag in dit verband is of deze ondernemingen beter presteerden dan zonder steun het geval zou zijn geweest - niet of zij beter presteerden dan niet-begunstigden in andere regio's.

Hetzelfde geldt voor algemene sectorale trends. Ook deze moeten worden meegewogen wanneer het effect van de steun moet worden geïdentificeerd. Zelfs indien begunstigden van regionale steun hun personeelsbestand inkrimpen, kan de steun toch nog doeltreffend zijn. Wanneer bijvoorbeeld de situatie in een bepaalde bedrijfstak als geheel verslechtert en alle ondernemingen banen schrappen, krimpen begunstigden van de steun hun personeelsbestand misschien minder sterk in dan zij anders hadden gedaan. Dit blijkt uit de onderstaande grafiek, die een negatieve trend laat zien in het aantal banen bij ondernemingen die steun ontvangen, zowel vóór als ná de toekenning van steun. Niettemin wordt deze trend minder negatief nadat de onderneming de steun heeft ontvangen. Wordt de trendlijn zonder steun en de trendlijn mét de reële werkgelegenheid die de onderneming aanbiedt nadat zij de steun heeft ontvangen, doorgetrokken, dan komt de positieve invloed van de steun in beeld.



**Figuur 1 - Positieve invloed van de steun bij een actuele negatieve trend**

Bij het identificeren van een controlegroep doet zich een specifiek probleem voor wanneer niet-begunstigden zelf hebben beslist of zij al dan niet steun aanvragen. Indien bijvoorbeeld alle ondernemingen subsidiabel zijn (d.w.z. alle ondernemingen die een project voorstellen en steun aanvragen, ontvangen ook bepaalde steun), dan zijn de ondernemingen die geen aanvraag indienen, waarschijnlijk de ondernemingen zonder project. Uit de resultaten voor de ondernemingen kan dan blijken dat ondernemingen die geen steun ontvingen, in absolute en in relatieve termen slechter presteerden dan de ondernemingen die wél steun ontvingen. Deze conclusie kan echter volledig worden verklaard door het enkele feit dat de eerste groep sowieso geen project had, terwijl dat bij de tweede groep wel het geval was, d.w.z. dat het in

de eerste groep het management aan belangstelling of creativiteit ontbrak. Daarom is het van cruciaal belang dat ondernemingen in de controlegroep (ondernemingen die geen steun genoten) in die groep zijn opgenomen om redenen die geen invloed hebben op de gemeten uitkomsten. Met name wanneer ondernemingen aan zelfselectie doen en er zelf voor gekozen hebben om geen steun aan te vragen, is mogelijk niet aan deze voorwaarde voldaan.

Systematische verschillen tussen begunstigden van staatssteun en niet-begunstigden dienen te worden ondervangen in de vormgeving van de evaluatie, zodat onzuiverheden in de resultaten worden voorkomen ("selectie-effect"). De afgelopen decennia zijn meerdere betrouwbare methoden ontwikkeld die voor deze problematiek een oplossing bieden. De keuze van de methode hangt af van de vormgeving van een bepaalde staatssteunregeling en van de beschikbare data. Elk van deze methoden kent haar beperkingen en zij zijn alleen valide wanneer bepaalde aannames stand houden. Deze beperkingen en aannames erkennen en openlijk bespreekbaar maken, is van cruciaal belang voor de geloofwaardigheid van een studie.

Een gerandomiseerde procedure voor de selectie van begunstigden is één middel om te voorkomen dat de evaluatie onzuiver is. Indien de selectie van begunstigden van de steun volledig gerandomiseerd gebeurt, kunnen eventuele systematische verschillen die bij de prestaties van de onderneming worden vastgesteld, aan de steun worden toegeschreven. Deze methode kan echter moeilijk praktisch toepasbaar zijn, met name voor grote bestaande regelingen. Andere methoden proberen bestaande factoren van exogene variatie in de omgeving waarbinnen ondernemingen opereren (d.w.z. variatie die niet wordt bepaald door parameters en variabelen in het model), te gebruiken om causale effecten te identificeren.<sup>10</sup> In bijlage I bij dit document worden de meest relevante methoden nader gepresenteerd, waarbij de klemtoon ligt op praktische aspecten van de toepassing ervan. Daar wordt ingegaan op de vraag hoe elk van deze methoden de causale effecten identificeert, omdat dit van bijzonder belang is in het kader van evaluaties van staatssteun waar de vooraf uitgetekende evaluatie ervoor moet zorgen dat een correcte evaluatie van de effecten van de steun mogelijk is.

Ten slotte dient te worden gecontroleerd op de impact van meervoudige steun (van een regeling, van meerdere regelingen of ad-hocsteun). Indien niet-begunstigden in een bepaald programma wel steun ontvangen uit andere programma's of indien begunstigden van een bepaald programma ook nog steun ontvangen uit andere programma's, zal de evaluatie van de effecten van dat bepaalde steunprogramma waarschijnlijk vertekend zijn.

### **3.5 Verzamelen van data: de best mogelijke bronnen gebruiken**

Consistente en voldoende hoeveelheden data moeten worden verzameld over zowel de begunstigden van steun als bij de controlegroep. Bepalen welke data nodig zijn, en toegang krijgen tot de bronnen van deze data is een onderdeel van de planning van de evaluatie.

---

<sup>10</sup> De meest gebruikte methoden zijn die van difference-in-difference, regression discontinuity, en instrumentele variabelen.

Daadwerkelijke monitoring van de interventie en accurate dataverzameling en -verwerking zijn van cruciaal belang om de kwaliteit van de evaluatie te borgen. Zodra de steunregeling goedkeuring heeft gekregen, zou er dus een mechanisme moeten worden uitgewerkt om de interventie te monitoren en om de nodige data te verzamelen en te verwerken. Hiermee zullen de kosten van de evaluatie waarschijnlijk fors verminderen.

Ervoor zorgen dat de nodige data over aanvragers en ontvangers van steun wordt verzameld, is een cruciale stap bij het uitwerken van het evaluatieplan, indien de beschikbaarheid van deze data een van de subsidiabiliteitsvoorwaarden kan worden.

Afgezien van data over steunaanvragen (met inbegrip van afgewezen aanvragers (voor zover beschikbaar)), moeten de databronnen voor begunstigden van de steun en voor de controlegroep identiek zijn, willen de data vergelijkbaar zijn. De kans is erg groot dat data uit meerdere bronnen zullen worden geput, door bijvoorbeeld data uit databases met gegevens over ontvangers van steun te combineren met data uit handelsregisters. De evaluatie moet misschien putten uit bestaande databronnen, zoals administratie databronnen (bijv. de belastingdienst, het handelsregister, innovatiesurveys en het octrooibureau). Daarom dient het evaluatieplan de bestaande databronnen te screenen, te beslissen of deze voldoende informatie bieden voor de evaluatie en ervoor te zorgen dat deze bronnen toegankelijk zullen zijn binnen het betrokken tijds kader.

Data uit administratieve bronnen (zoals nationale statistische bureaus) zullen voor de evaluatoren waarschijnlijk alleen beschikbaar zijn onder bepaalde voorwaarden met betrekking tot privacy en vertrouwelijkheid van bedrijfsgegevens. De voorwaarden om toegang te krijgen tot deze data moeten in het evaluatieplan worden beschreven. Wanneer dat nodig is, moet de autoriteit die toegang tot de data verleent, ervoor zorgen dat de deskundigen die de evaluatie verrichten, toegang hebben tot deze data.

Bij het gebruik van data uit verschillende bronnen is het van bijzonder belang dat deze data worden verzameld in een formaat dat een consistente matching van variabelen mogelijk maakt. Misschien moeten voor observatie-eenheden unieke identificatiecodes worden gezocht die in elke dataset kunnen worden gebruikt. Zo moeten identificatoren voor bedrijven en bedrijfslocaties uniek zijn in alle datasets, moeten adressen worden verzameld in een formaat dat geolokalisatie mogelijk maakt enz. De precieze herkomst van de identifier kan verschillen afhankelijk van de lidstaat. Deze kan bijvoorbeeld afkomstig zijn van de belastingdienst (bijv. een btw-nummer) of rechtstreeks zijn meegedeeld door statistische instellingen (bijv. het bedrijfsidentificatienummer SIREN en het vestigingsnummer SIRET die door het Franse bureau voor de statistiek INSEE worden toegekend).

De evaluatie van staatssteun kan worden aangevuld met gegevens uit enquêtes van begunstigden van de steun en/of interviews met beheerders van regelingen. Dit soort kwalitatieve informatie is, naar haar aard, subjectief en in antwoorden kunnen de strategische belangen van de begunstigden doorklinken, in plaats van dat zij een reële inschatting geven van het effect van de steun. Dit risico is bijzonder groot indien de geënquêteerde in de overtuiging verkeert dat een positief getuigenis zijn kansen vergroot om in de toekomst steun

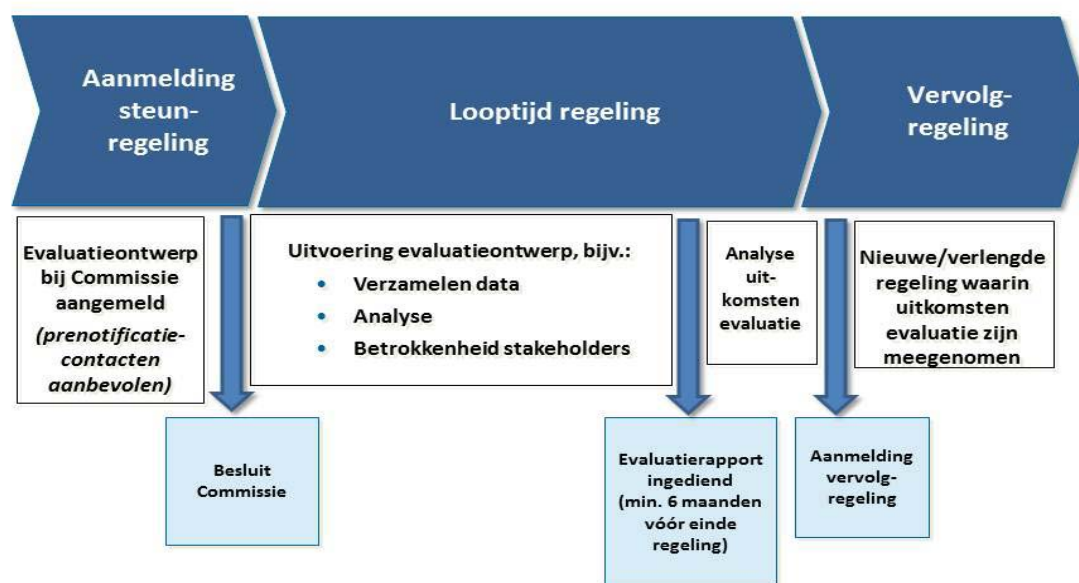
te krijgen. Niettemin kan informatie uit kwalitatieve exercities zoals interviews en casestudies, mits deze met de nodige mate van voorzichtigheid wordt gehanteerd, een nuttige, aanvullende bron zijn en kan zij helpen bij de interpretatie van de resultaten van de evaluatie.

Wanneer bij evaluaties persoonsgegevens worden verwerkt, is de EU-wetgeving betreffende de bescherming van persoonsgegevens van toepassing, in het bijzonder Richtlijn 95/46/EG betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens en de desbetreffende nationale uitvoeringsmaatregelen, alsmede Verordening (EG) nr. 45/2001 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door de communautaire instellingen en organen en betreffende het vrije verkeer van die gegevens.

### **3.6 Tijdpad evaluatie**

Een evaluatieplan moet informatie bevatten over het precieze tijdpad van de evaluatie. Dat zal worden bepaald overeenkomstig de goedgekeurde looptijd van de regeling en moet mijlpalen bevatten - voor het verzamelen van data, voor het uitvoeren van de evaluatie en voor het indienen van het eindrapport. Het tijdpad kan verschillen afhankelijk van de regeling en dient dus van geval tot geval te worden overlegd en overeengekomen met de Commissie. Beheerders van regelingen zouden zich best openstellen voor informeel overleg over de inhoud van het plan voordat zij hun officiële aanmelding bij de Commissie indienen.

Om een beoordeling te kunnen maken van een voorgenomen verlenging van een steunregeling, dient het evaluatierapport voldoende op tijd bij de Commissie te worden ingediend (bijv. zes maanden vóór het geplande einde van de regeling). Wordt geen verlenging van de regeling overwogen, dan kan het rapport worden ingediend zodra de regeling is afgelopen.



**Figuur 2 - Overzicht van het evaluatieproces in het geval van een aangemelde regeling**

### **3.7 De evaluerende instantie: onafhankelijkheid en deskundigheid verzekeren**

De evaluatie van de effecten van staatssteunregelingen dient objectief, streng, onpartijdig en transparant te gebeuren.<sup>11</sup> Iedere evaluatie dient op basis van degelijke methoden te worden uitgevoerd door deskundigen die beschikken over afdoende en bewezen deskundigheid en de methodologische kennis om de exercitie te kunnen uitvoeren.

Evaluaties dienen te worden uitgevoerd door een instantie die ten minste functioneel onafhankelijk is van de steunverlenende autoriteit en die beschikt over de nodige en bewezen knowhow en voldoende gekwalificeerde medewerkers om dit soort evaluaties uit te voeren. De eis dat de evaluator functioneel onafhankelijk is van de steunverlenende autoriteit, is van cruciaal belang om de kwaliteit en de geloofwaardigheid van de evaluatie te kunnen borgen. Dit betekent niet noodzakelijk dat een nieuwe instantie moet worden opgericht, noch dat de evaluatiebehoefte aan commerciële evaluatoren moet worden uitbesteed. Afhankelijk van de specifieke organisatievormen in elke lidstaat is het bijvoorbeeld mogelijk om een beroep te doen op de onafhankelijkheid en de knowhow van organisaties zoals statistische bureaus, centrale banken, rekenkamers, publieke of particuliere universiteiten of onderzoekscentra. Hierover kan individueel per regeling worden beslist.

<sup>11</sup> Zie bijv. de *Evaluation Standards* van de Europese Commissie, de *Evaluation Norms and Standards* van de OESO, de *Evaluation Standards* van de Verenigde Naties en de *Independent Evaluation: Principles, Guidelines and Good Practice* van de Wereldbank.

Voor het welslagen van een evaluatie is het van belang om de evaluerende instantie in een vroeg stadium (bijv. tijdens het uitwerken van de regeling) te betrekken. Dit zorgt ervoor dat de steunregeling zal kunnen worden geëvalueerd zoals dat was bedoeld, en dat de nodige garanties voorhanden zijn dat de vereiste data zullen worden verzameld. Voor zover mogelijk moet het evaluatieplan steeds worden opgesteld door (of zeker in bijzonder nauwe samenwerking met) de aangewezen evaluator. Het moet ook informatie bevatten (ook al zijn dat maar indicaties) over de vereiste personele en financiële hulpmiddelen die beschikbaar zullen worden gesteld om de evaluatie uit te voeren. Informatie over de identiteit en de rol van iedere cruciale deskundige die bij de evaluatie betrokken is en een inschatting van zijn aandeel daaraan, zijn daarbij van bijzonder belang.

Het evaluatieplan moet een precieze beschrijving geven van de evaluerende instantie of, indien die nog niet is geselecteerd, de gedetailleerde criteria die voor de selectie van die instantie zullen worden gehanteerd, met name wat betreft onafhankelijkheid, deskundigheid en knowhow. Het moet, voor zover mogelijk, ook alternatieven geven. Wanneer de evaluerende instantie nog niet is geselecteerd (of reeds is geselecteerd, maar nog niet actief bij het opstellen van het evaluatieplan betrokken was), moet duidelijk worden aangegeven waarom dat het geval was. Zelfs in dit soort situatie moet het evaluatieplan voldoende gedetailleerd zijn om een correcte beoordeling te kunnen maken van de validiteit van de uit te voeren evaluatie.

### **3.8 Ruchtbaarheid geven aan de evaluatie: de betrokkenheid van stakeholders stimuleren**

De evaluatie moet worden openbaargemaakt. Dit betekent dat aan zowel het evaluatieplan als het definitieve evaluatierapport, na goedkeuring daarvan, voldoende ruchtbaarheid moet worden gegeven door het beschikbaar te stellen op de in het evaluatieplan beschreven locaties, zoals bijvoorbeeld een website. Ook de Commissie zou die documenten kunnen openbaar maken.<sup>12</sup>

Zijn de voor de evaluatie gebruikte data persoonsgegevens en/of vertrouwelijke informatie, dan dient de vertrouwelijkheid te worden gegarandeerd gedurende het hele evaluatieproces, met name in overeenstemming met de artikelen 8, 16 en 17 van het Handvest van de grondrechten van de Europese Unie. Niettemin geldt de vertrouwelijkheid niet voor de resultaten van de evaluatie. Met name kan in het contract voor de evaluatie geen vertrouwelijkheidsclausule worden opgenomen, afgezien van: i) geheimhoudingsverplichtingen voor persoonsgegevens en/of vertrouwelijke gegevens, en ii) verplichtingen om de algemene bepalingen in acht te nemen van de nationale wetgeving op de

---

12 Met uitzondering van bedrijfsgevoelige en anderszins vertrouwelijke informatie in goed onderbouwde gevallen; zie de mededeling van de Commissie C(2003) 4582 van 1 december 2003 over geheimhouding bij beschikkingen inzake staatssteun (PB C 297 van 9.12.2003, blz. 6). Wanneer persoonsgegevens worden openbaar gemaakt, moet dit gebeuren in overeenstemming met het Unierecht betreffende de bescherming van persoonsgegevens, met name Richtlijn 95/46/EG en de desbetreffende nationale uitvoeringsmaatregelen, alsmede Verordening (EG) nr. 45/2001.

statistiek en van de statistische geheimhouding, onder meer wat betreft de presentatie van de resultaten.

De bij de evaluatie verzamelde data moeten, met het oog op het repliceren van resultaten of met het oog op verdere studie, beschikbaar worden gesteld op voorwaarden die niet restrictiever zijn dan de voorwaarden die werden opgelegd aan de instantie die de initiële evaluatie uitvoerde.

De steunverlenende autoriteit zou ervoor kunnen zorgen dat relevante stakeholders bij de exercitie betrokken worden; zij zouden tijdens de tenuitvoerlegging van het evaluatieplan ten minste eenmaal moeten worden geconsulteerd. Stakeholders zou bijvoorbeeld kunnen worden gevraagd om, op basis van een tussentijds rapport, te overleggen over de initiële bevindingen van de evaluatie. Dit soort regelingen zou in het evaluatieplan moeten worden opgenomen.

#### **4 Criteria voor de selectie van te evalueren steunregelingen**

In beginsel komt iedere staatssteunregeling voor evaluatie in aanmerking. Maar ook al wordt evaluatie als een goede praktijk beschouwd, toch is deze volgens de staatssteunregels niet in alle gevallen vereist. Evaluatie van staatssteun zou een evenredige exercitie moeten blijven en zou, algemeen genomen, moeten worden uitgevoerd voor regelingen die potentieel een aanzienlijk effect op de interne markt kunnen hebben en een risico in zich dragen op aanzienlijke verstoring, indien de tenuitvoerlegging niet tijdig wordt doorgelicht. In de desbetreffende staatssteunrichtsnoeren ligt daarom de klemtoon op: i) grote steunregelingen, ook wanneer deze onder de algemene groepsvrijstellingsverordening vallen; ii) vernieuwende steunregelingen, of iii) steunregelingen waarbij voor de nabije toekomst moet worden gerekend met mogelijk ingrijpende veranderingen (wat betreft markten, technologie of regelgeving) waardoor de beoordeling van de regeling wellicht opnieuw moet worden bezien. In de individuele staatssteunrichtsnoeren worden ook andere soorten regelingen genoemd die bij een evaluatie baat zouden vinden.

##### **4.1 Grote steunregelingen, ook steunregelingen in het kader van de algemene groepsvrijstellingsverordening**

Aansluitend bij de mededeling van de Commissie over de modernisering van het EU-staatssteunbeleid (SAM) zou de Commissie kunnen eisen dat de grootste steunregelingen worden geëvalueerd, omdat: i) dit soort regelingen een bijzonder sterke impact kunnen hebben op de interne markt indien zij niet goed zijn vormgegeven; ii) daarmee de grootste efficiencywinsten kunnen worden geboekt dankzij de grote budgetten ervan, en iii) grote regelingen met vele uiteenlopende soorten begunstigden voldoende data kunnen opleveren met het oog op evaluatie.

Voor bepaalde steunregelingen hoeft misschien toch nog geen evaluatie te worden uitgevoerd indien deze, ondanks de omvang ervan, geen specifieke problematische aspecten opleveren (bijv. routinezaken, gevallen waarin een groot aantal begunstigden telkens kleine bedragen aan steun ontvangt, en zaken die geen risico inhouden op ingrijpende veranderingen of waar geen ernstige verstoringen kunnen ontstaan).



Daarnaast definieert de nieuwe algemene groepsvrijstellingsverordening (AGV) grote steunregelingen op basis van het budget ervan (een gemiddeld jaarbudget van meer dan 150 miljoen EUR) en eist zij dat voor bepaalde categorieën steun<sup>13</sup> een evaluatie plaatsvindt.

Om de inwerkingtreding van deze grote regelingen niet te vertragen, maar ook om te verzekeren dat zij daadwerkelijk worden geëvalueerd, wordt in de algemene groepsvrijstellingsverordening voorzien in een afwijking van de aanmeldingsverplichting voor een periode van maximaal zes maanden (door de Commissie eventueel te verlengen na goedkeuring van het evaluatieplan)<sup>14</sup>. Het evaluatieplan moet zo snel mogelijk worden aangemeld, doch uiterlijk binnen 20 werkdagen na de inwerkingtreding van de regeling.

Volgens de nieuwe algemene groepsvrijstellingsverordening moet ook bij aanpassingen aan of vervolgregelingen van grote regelingen een evaluatie plaatsvinden, tenzij het om louter formele of administratieve aanpassingen gaat of het aanpassingen betreft die plaatsvinden in het kader van maatregelen met EU-cofinanciering.

#### **4.2 Vernieuwende steunregelingen**

Wat als "vernieuwend" geldt, kan variëren afhankelijk van de steuninstrumenten en van lidstaat tot lidstaat. Een steunregeling geldt in beginsel als vernieuwend in termen van de aard van de regeling of de markten waarop zij is gericht, bijv. opkomende markten waar marktontwikkelingen nog in een zeer vroeg stadium verkeren. Dit soort regelingen dragen het potentieel in zich dat zij blijvend en fundamenteel het uitzicht van een bedrijfstak kunnen bepalen. Daarom zijn de mogelijkheden voor voordelen - en verstoringen - bijzonder groot. Bij dit soort vernieuwingen kan het bijvoorbeeld gaan om de introductie van een nieuw capaciteitsmechanisme in de energiesector, steun voor nieuwe soorten technologie, of een vernieuwende vorm van steun voor hernieuwbare energiebronnen in het kader van milieusteun. De evaluatie van vernieuwende regelingen helpt ook bij de vormgeving van nieuwe regelingen, omdat zo de laatste ontwikkelingen op de markt kunnen worden meegenomen.

#### **4.3 Steunregelingen beïnvloed door te verwachten ingrijpende veranderingen**

De kans op ingrijpende veranderingen (wat betreft markten, technologie of regelgeving) in de nabije toekomst zal van geval tot geval worden beoordeeld. Bij dit soort ingrijpende veranderingen kan het bijvoorbeeld gaan om de versnelde herziening van geldende regelgeving of steun voor snel evoluerende bedrijfstakken waar de marktomgeving en de beschikbare technologie in een razendsnel tempo verlopen. Indien die regelingen niet worden aangepast aan de effecten van die ingrijpende veranderingen, bestaat het risico dat overheidsmiddelen niet doelmatig zullen worden ingezet (bijvoorbeeld omdat financiering kan gaan naar potentieel "marktfalen" dat zal verdwijnen), of dat aanzienlijke verstoringen

---

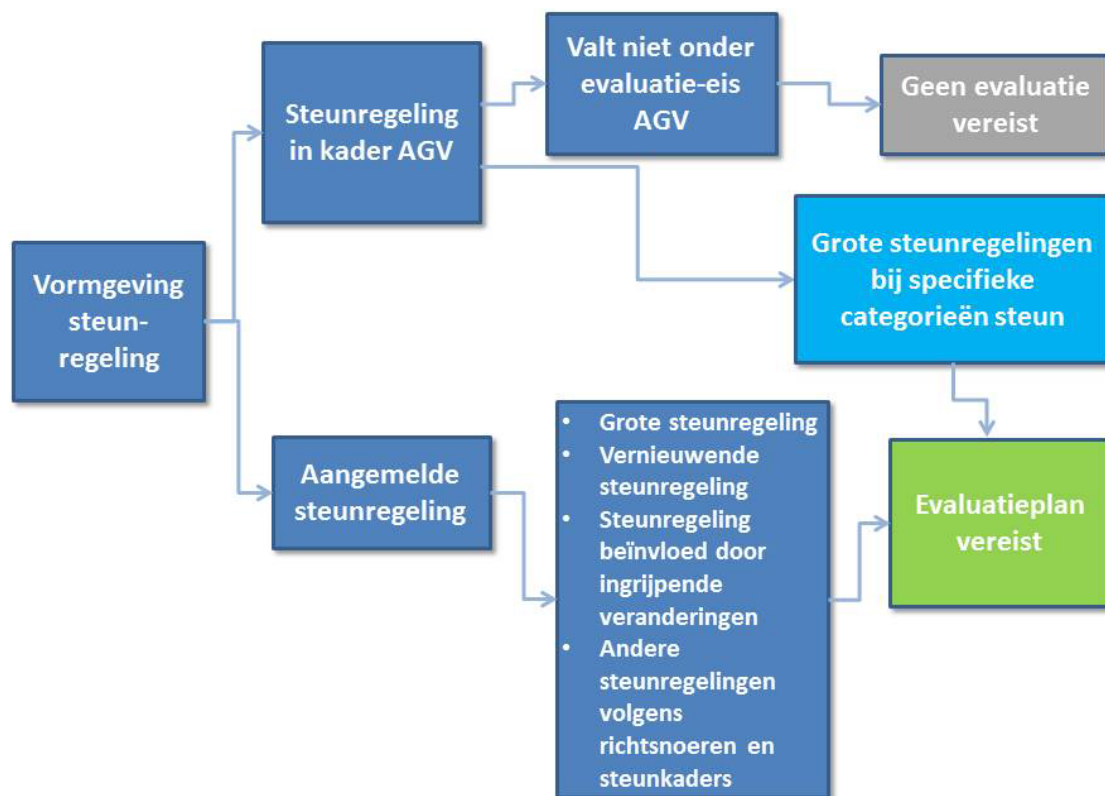
13 Regionale steun (met uitzondering van regionale exploitatiesteun), mkb-steun, steun om het mkb toegang te geven tot financiering, O&O&I-steun, milieusteun (met uitzondering van steun in de vorm van kortingen op milieubelastingen op grond van Richtlijn 2003/96/EG) en steun voor breedbandinfrastructuur

14 De Commissie zou ook kunnen beslissen dat, bij wijze van uitzondering, geen evaluatie vereist is, gelet op de specifieke kenmerken van de zaak.

zullen optreden die nieuwkomers op de markt anders zullen treffen dan gevestigde ondernemingen of dat ongelijke voorwaarden ontstaan voor nieuwe technologie en legacy-technologie. Treffende voorbeelden in dit verband zijn de herziening van een bestaand reguleringskader (bijv. in de sector elektronische communicatie), de sterke fluctuatie van in- of outputprijzen (bijv. in het geval van zonnepanelen) of de introductie van een nieuwe technologie op de markt (bijv. de beschikbaarheid van een 4G-breedbandnetwerk). Telkens gaat het hier om gevallen waar evaluatie verantwoord zou zijn, zodat nieuwe marktontwikkelingen bij toekomstige regelingen worden meegenomen.

#### 4.4 Andere steunregelingen

In de richtsnoeren voor de verschillende takken van het staatssteunbeleid wordt ook een aantal steunregelingen genoemd waarvoor evaluatie bijzonder relevant zou zijn.



**Figuur 3 - Selectie steunregelingen voor evaluatiedoeleinden**

## **Bijlage I: Technische bijlage over relevante methoden om de causale effecten te identificeren**

Een staatssteunregeling kan op sterk uiteenlopende niveaus effecten hebben. Normaal gesproken is de verwachting dat de regeling een direct effect heeft op het niveau van de begunstigde. Inzicht krijgen in de omvang van dit effect is van cruciaal belang om na te gaan hoe doelmatig en doeltreffend een overheidsmaatregel is. Omdat steun echter is gericht op ondernemingen die interageren met markten of regio's die concurreren om economische activiteit aan te trekken, heeft staatssteun meestal ook indirecte effecten. Bij deze effecten kan het bijvoorbeeld gaan om overloopeffecten naar andere ondernemingen (bijv. positieve overloopeffecten van O&O of de verdringing van investeringen van andere, concurrerende ondernemingen) of verplaatsingseffecten (bijv. verschuivingen van economische activiteiten van een regio naar een andere). Deze indirecte effecten veroorzaken zowel potentiële schade als voordelen van overheidsinterventie in de economie. Daarom moet bij de evaluatie van overheidsmaatregelen ook de omvang van deze indirecte effecten worden gemeten.

Om de directe en indirecte effecten van een beleid te meten, moet doorgaans met verschillende instrumenten worden gewerkt. De afgelopen decennia hebben de methodieken en technieken voor het meten van het directe effect van beleidsmaatregelen op de begunstigden ervan zich sterk ontwikkeld. Deze technieken worden verderop in deze bijlage meer in detail gepresenteerd. Jammer genoeg zal het slechts zelden mogelijk zijn om met deze technieken ook de indirecte effecten van de steunregeling op ondernemingen of regio's te meten. Voor de evaluatie van de indirecte effecten van de staatssteunregeling zijn doorgaans andere soorten bewijsmateriaal vereist dan wat wordt gebruikt voor de meting van de directe effecten op de begunstigden van de steun. Ook bij de interpretatie wordt doorgaans meer gebruikgemaakt van economische theorieën en modellen. Het is veel moeilijker om voor dit soort exercitie specifieke houvast te bieden, omdat hier moet worden ingezoomd op de mogelijke en verwachte positieve en negatieve effecten van het beleid. Daarom dient deze evaluatie te worden uitgevoerd na een zorgvuldige en strenge analyse van de meest geloofwaardige indirecte effecten van het beleid. Op basis van deze analyse kunnen evaluatoren dan maatstaven afleiden die zijn gebaseerd op microdata van ondernemingen die geen steun ontvangen, met name in dezelfde regio, hetzelfde cluster of dezelfde bedrijfstak, maar ook in aangrenzende regio's. Dit zou de kern moeten zijn van de meting van de indirecte effecten van de staatssteunregeling. Indien nodig kan een en ander verder worden aangevuld met meer macro-economische data en - vooral - zorgvuldig geselecteerde gevalstudies.

De evaluatie van directe effecten is een noodzakelijke en cruciale eerste stap. Toch kan een zorgvuldig uitgevoerde meting van de indirecte effecten van de steun belangrijk bewijsmateriaal opleveren voor de meting van de ruimere effecten van de regeling. Mag het uitblijven van bijkomende investeringen door begunstigden van de steun, over het algemeen genomen, een aanwijzing zijn dat het beleid heeft gefaald, toch is zelfs een positief effect niet voldoende om te concluderen dat een beleidsmaatregel zijn doelstellingen heeft bereikt. Met name wanneer blijkt dat het directe effect van de steun op de begunstigden erg gering is of

zelfs onbestaande, is de kans groot dat de regeling wordt beschouwd als een regeling die haar doelstelling niet heeft waargemaakt, tenzij zeer overtuigende argumenten kunnen worden gegeven voor het bestaan van brede en gunstige indirecte effecten. Het tegendeel geldt evenzeer: indien de conclusie van de evaluatie is dat de steun positieve directe effecten heeft, blijft de vraag of er misschien negatieve indirecte effecten zijn die deze positieve effecten compenseren of daar zelfs tegen opwegen.

Bovendien is het niet altijd eenvoudig om een duidelijk onderscheid te maken tussen directe en indirecte effecten. Een onderneming heeft misschien meer geïnvesteerd (beweerd direct effect) omdat haar eigen investering investeringen van concurrerende ondernemingen heeft weggedrukt (interactief indirect effect). Een onderneming kan ook meer investeren omdat zij rekent met overloopeffecten van en investeringen door andere ondernemingen. Voorts kan ook de steun zelf of alleen al de steunverlening een van deze effecten hebben. Bij de evaluatie van de directe effecten moet ook in detail worden ingegaan op de vraag in hoeverre indirecte zullen spelen, in welke richting zij zullen gaan en wat daarvan de omvang zal zijn. De economische theorie die de indirecte effecten aan de steun koppelt, dient uitdrukkelijk te worden aangegeven. Ook dient verdere informatie die kan dienen als bewijsmateriaal ter staving van deze theorie, integrerend deel uit te maken van de evaluatie.<sup>15</sup>

### *Causale inferentie*

Het causale effect van steun is het verschil tussen de uitkomst mét steun en de uitkomst zónder steun. De uitkomst mét steun wordt geobserveerd voor ondernemingen die de steun ontvangen. De uitkomst zonder steun wordt echter alleen gemeten voor ondernemingen die geen steun ontvangen. Per definitie valt niet te observeren wat de uitkomst zonder steun was geweest voor de ondernemingen die wel steun hebben ontvangen. Om het effect van de steun op de begunstigde ondernemingen in te schatten, moet dus dit *nulscenario* worden geconstrueerd: een redelijk scenario moet worden uitgewerkt dat in beeld brengt wat waarschijnlijk met de ontvangers van de steun was gebeurd indien zij de steun níét hadden gekregen. Hiervoor moet een *controlegroep* worden bepaald, d.w.z. een groep ondernemingen die, afgezien van de steun zelf, voor alle aspecten zo goed mogelijk overeenkomt met de groep ondernemingen die de steun heeft gekregen.

De kwaliteit van deze controlegroep is van cruciaal belang voor de validiteit van de evaluatie. De kenmerken van ondernemingen die steun krijgen, verschillen doorgaans van ondernemingen die geen steun krijgen. Zo zijn zij bijvoorbeeld actief in een armer gebied met minder marktpotentieel, zijn hun kredietmogelijkheden beperkter, zijn zij meer of minder doelmatig, hebben zij wel of niet een project om uit te voeren enz. Daarom zal een naïeve vergelijking van ondernemingen die wel steun krijgen met ondernemingen die geen steun ontvangen, waarschijnlijk meer deze realiteit tot uiting brengen dan het effect van de steun zelf.

---

15 In dit werkdocument ligt de klemtoon op de directe effecten van steun. Toch zorgt het feit dat steun indirecte effecten kan hebben, voor een aantal analytische uitdagingen met betrekking tot de meting van directe effecten. Daarbij dient bijzondere aandacht te gaan naar de effecten van interactie met de markt.

Om een valide evaluatie te kunnen uitvoeren, komt het er dus op aan ervoor te zorgen dat dit systematische verschil tussen ondernemingen die wel en ondernemingen die geen staatssteun ontvangen (het zogeheten "selectie-effect"), geen onzuivere resultaten oplevert. De afgelopen decennia zijn diverse betrouwbare methoden ontwikkeld om voor deze problematiek een oplossing te bieden. De keuze van de methode is afhankelijk van het te evalueren beleid en de beschikbare data. Bovendien heeft elk van deze methoden zijn beperkingen en heeft zij alleen voor een aantal aannames validiteit. De geloofwaardigheid van een studie kan worden vergroot door deze beperkingen expliciet te benoemen en te bespreken. In deze technische bijlage worden de meest relevante methoden gepresenteerd. De aandacht gaat vooral naar de meest praktische aspecten en de klemtoon ligt op het belang van een goede identificatiestrategie.<sup>16</sup>

### **A. Gerandomiseerde experimenten**

Het bepalen van een goede controlegroep is van essentieel belang om goede (zuivere) metingen te krijgen van het beleidseffect. Het gunstigste geval is dat waarin er geen selectie-effect speelt omdat de begunstigten gerandomiseerd zijn geselecteerd.<sup>17</sup> In dat geval is er, afgezien van de steun, geen systematisch verschil tussen begunstigten en niet-begunstigten en kan het verschil in de uitkomsten aan het beleid worden toegewezen.

Toch komt er soms kritiek op een gerandomiseerde selectie van begunstigten van steun omdat deze zou indruisen tegen de doelstelling van vele regelingen om, op basis van objectieve criteria, net de best mogelijke begunstigten van de steun te selecteren. Niettemin is het in bepaalde omstandigheden misschien toch mogelijk om voor de subsidiabiliteit of de prikkels voor begunstigten om deel te nemen, elementen van randomisatie te introduceren. Een voorbeeld daarvan is het bepalen van een vast budget voor een bepaalde regeling. Indien de steunverzoeken van de aanvragers het budget overtreffen en zij vrij gelijklopende kenmerken hebben, kan worden geprobeerd om treatments te randomiseren. Een ander voorbeeld is dat kandidaat-steunontvangers op aselechte basis met verschillende niveaus van informatie over de regeling te maken krijgen.

Pilotprojecten bieden verdere kansen voor gerandomiseerde toewijzing van steun. In het geval van innovatief beleid verdient het misschien aanbeveling om eerst een kleinschaligere pilot te evalueren. Deze pilot kan beperkter in omvang zijn en begunstigten kunnen dan gemakkelijker gerandomiseerd worden geselecteerd. Een ander alternatief is het infaseren van een regeling, waarbij bijvoorbeeld 25% van gerandomiseerd geselecteerde ondernemingen in aanmerking komt in jaar 1, 50% in jaar 2, 75% in jaar 3 en 100% in jaar 4. (Ook kan worden overwogen om de regeling bij een steeds ruimer publiek bekend te maken.) Bij nieuw beleid is een infaseringsperiode vaak ook een administratieve noodzaak.

---

16 Deze bijlage biedt een snelle en niet-technische presentatie van de econometrische methoden voor beleidsevaluatie. Voor deze presentatie is gebruikgemaakt van vele elementen uit Givord (2010); andere zeer goede presentaties zijn te vinden bij Imbens en Wooldridge (2009) en Angrist en Pischke (2008).

17 Gerandomiseerde experimenten waren bijvoorbeeld gedurende tientallen jaren de enige acceptabele methode voor de meting van de effecten van geneesmiddelen en medische behandelingen (vandaar ook het gebruik van de term "treatment").

Deze ideeën zijn misschien beter bruikbaar voor de tenuitvoerlegging van volledig nieuwe regelingen of een ingrijpende verandering van bestaande regelingen. Het is waarschijnlijk vrij moeilijk om de subsidiabiliteit (direct of indirect) te randomiseren bij de voortzetting van een bestaande regeling. Toch betekent dit niet dat gerandomiseerde experimenten niet zouden kunnen worden gebruikt voor delen van de evaluatie ervan. Met name is het nog steeds mogelijk om begunstigden gerandomiseerd te selecteren op potentieel doelmatigere, doelgerichtere en/of minder verstorende varianten van de regeling. Zo kan het bijvoorbeeld in het geval van een subsidierегeling mogelijk zijn om, in plaats daarvan, een nieuw opgezette leningregeling gerandomiseerd aan te bieden.

## **B. Quasi-experimentele methoden**

Ook al zijn gerandomiseerde experimenten de best mogelijke methode om het effect van beleidsmaatregelen te evalueren, toch is het niet steeds mogelijk om deze toe te passen. Andere methoden zijn ontwikkeld om de effecten van een beleid te evalueren uit een ex-post-perspectief. Zij delen dezelfde doelstelling om exogene variaties van de omgeving waarbinnen ondernemingen opereren, te gebruiken voor het creëren van situaties die erg nauw aansluiten bij experimenten (de zogeheten natuurlijke of quasi-experimenten).

Het is doorgaans een uitdaging om, met het oog op een ex-post-beoordeling, natuurlijke of quasi-experimenten uit te werken. Toch kan een nauwgezette analyse van de vormgeving van het beleid leiden tot de bevinding dat er voldoende exogene variatie bestaat. Zo nodig kan het initiële opzet worden aangepast zodat meer elementen worden opgenomen om de effecten van het beleid te identificeren.

### *Controle op observeerbare verschillen*

Normaal gesproken bestaan er, zoals gezegd, aanzienlijke verschillen tussen ondernemingen die wél en ondernemingen die geen steun ontvangen. Daarom moeten die verschillen kunnen worden verklaard wanneer een vergelijking wordt gemaakt van de uitkomsten voor de groep van de ondernemingen mét steun en die zonder steun.

Veel van die verschillende kenmerken zijn meestal observeerbaar. Vaak wordt voor het verklaren van die verschillen gebruikgemaakt van een lineair regressiemodel. Bij lineaire regressie wordt geprobeerd te controleren op de invloed van geobserveerde kenmerken op de uitkomsten. De aanname daarbij is dat er een lineair verband bestaat tussen de uitkomst (bijv. de O&O-investering) en andere kenmerken van de onderneming (bijv. de bedrijfstak, leeftijd, omvang enz., met inbegrip van de steunverlening). Lineaire regressie kan worden gezien als een lineaire benadering van complexere verbanden.<sup>18</sup> Lineaire regressies kunnen worden gezien als technieken voor algemeen gebruik en vinden toepassingen in vele uiteenlopende evaluatieprocessen.

---

18 Bovendien bieden zij de mogelijkheid tot interactie van kenmerken (bijv. omzet en sector) en om functies aan deze kenmerken toe te kennen (bijv. kwadraten van variabelen).

Een alternatief voor lineaire regressie is het gebruik van matchingtechnieken. Matchingtechnieken zijn bedoeld om iedere begunstigde te koppelen aan een andere onderneming die er zeer gelijkend "uitziet", maar die geen steun heeft ontvangen. Bij matching kan voor de observeerbare kenmerken worden gebruikgemaakt van de kenmerken van de onderneming of de geschatte waarschijnlijkheid dat zij steun ontvangt (de zgn. *propensity score matching*). Matching kan een nuttige methode zijn om voor observeerbare factoren te controleren in het kader van een valide empirische strategie.

Toch zijn aan zowel enkelvoudige lineaire regressie als matching een aantal intrinsieke beperkingen verbonden. Beide methoden hebben alleen validiteit in de aanname van de zogeheten conditionele onafhankelijkheid. Dit betekent dat, zodra de effecten van de observeerbare kenmerken in aanmerking zijn genomen, de uitkomst onafhankelijk is van de observeerbare kenmerken. In de praktijk is daarvoor normaal gesproken vereist dat iedere variabele die een effect heeft op zowel de uitkomst als de selectie, observeerbaar is (en met de juiste functionele vorm wordt meegewogen). Indien dat niet het geval is, kan het enkele feit dat een onderneming deelneemt, een weerspiegeling zijn van bepaalde (niet-geobserveerde) kenmerken van die onderneming die haar prestaties evenzeer bepalen. Zowel lineaire regressie als matching zullen geen valide evaluatie opleveren. Wanneer een onderneming bijvoorbeeld een "veelbelovend project" heeft, heeft dit zowel invloed op de waarschijnlijkheidsgraad dat zij steun zal aanvragen (én krijgen) als de waarschijnlijkheid dat de onderneming erin slaagt om haar activiteiten te laten groeien. Deze elementen níet laten meewegen zal onzuivere resultaten opleveren.

Meer bepaald kan wanneer bij matching een begunstigde wordt vergeleken met zijn gematchede "tweeling" zonder steun, het selectie-effect alleen worden vermeden indien de toekenning van de steun geen verband houdt met niet-geobserveerde variabelen die de uitkomsten ook beïnvloeden. In werkelijkheid zal deze aanname zelden vervuld zijn. Alle variabelen meten die een effect hebben op het aanvragen of ontvangen van steun, is zelden mogelijk. Om matchingtechnieken te kunnen toepassen, is bovendien vereist dat ondernemingen die steun ontvangen, in hun observeerbare kenmerken sterk gelijken op ondernemingen die geen steun ontvangen. Indien de gematchede ondernemingen werkelijk voor ieder observeerbaar aspect vergelijkbaar zijn, wordt de reden waarom bepaalde ondernemingen steun krijgen en andere niet, per definitie niet geobserveerd. De verantwoording van de validiteit van een evaluatie op basis van matching of een eenvoudiger, traditionele lineaire regressieanalyse kan dus niet gewoonweg zijn dat er een zeer volledige dataset met talrijke geobserveerde kenmerken bestaat.

Integendeel, de potentiële argumenten om bij evaluatie matching of de eenvoudiger lineaire regressie te gebruiken, zijn gebaseerd op het feit dat deze niet-geobserveerde redenen die de subsidiabiliteit of de toekenning van steun verklaren, geen directe of indirecte invloed hebben op de uitkomsten (zodra deze zijn gecontroleerd op observeerbare factoren). Wil een evaluatie op basis van eenvoudige matching of lineaire regressie validiteit hebben, dan moet er zekerheid bestaan dat de reeks ondernemingen die geen steun hebben ontvangen, exogeen is bepaald. Daartoe is vereist dat, zodra is gecontroleerd op de observeerbare factoren, er geen

niet-geobserveerde factor overblijft die de subsidiabiliteit of de toekenning van de steun verklaart en die ook direct of indirect invloed zou hebben op de uitkomsten. Over het algemeen zal matching van ondernemingen die ook subsidiabel zijn, niet aan dit laatste criterium voldoen. Indien bijvoorbeeld alle ondernemingen subsidiabel zijn, is de waarschijnlijkheid groter dat ondernemingen die investeringssteun krijgen, een project hebben dan ondernemingen die geen steun krijgen (omdat zij anders ook een aanvraag hadden ingediend en steun hadden gekregen). Al bij al is de kans groter dat ondernemingen mét een project groeien in termen van omzet of banen, maar houdt een en ander geen verband met steun en is het met matching op basis van observeerbare factoren niet mogelijk beide te ontwarren (tenzij het bestaan van een vergelijkbaar investeringsproject wordt gemeten).

In vele situaties is de aanname van de conditionele onafhankelijkheid wel tot mislukken gedoemd. Daarom kan het nodig zijn om andere technieken toe passen dan gewone lineaire regressie of matching, om bij het treatment het bestaan van niet-geobserveerde selectie te verklaren.

De rest van deze bijlage presenteert de meest gebruikte methodieken om binnen deze context de beleidsimpact meer in detail te meten: differences-in-differences, regression discontinuity design, instrumentele variabelen. Deze methodieken ontleen hun validiteit aan andere aannames. Wat daarbij de beste keuze is, wordt meestal bepaald door de beleidscontext en de beschikbaarheid van data. Bij deze presentatie worden de sterke en zwakke punten van elk van deze technieken besproken. Afgezien van de opvallende uitzondering van gerandomiseerd onderzoek met controlegroep (RCT) die hier reeds aan bod kwam, is geen van deze technieken voor ieder aspect superieur aan alle andere. De keuze voor een bepaalde techniek moet voortvloeien uit een zorgvuldige analyse van de context van de maatregel en de beschikbare data.

In dit verband is het niet zonder belang te beklemtonen dat met een bepaalde econometrische techniek niet de effecten van een bepaald beleid kunnen worden geïdentificeerd, maar wel de exogeniteit van de controlegroep - en dus de kwaliteit van het nulscenario. Voor de kwaliteit van de evaluatiestudie zal dus bepalend zijn hoe overtuigend de onderzoeker de exogeniteit van de controlegroep kan vaststellen. In gevallen waar er mogelijk nog storingsfactoren overblijven, is het van essentieel belang om deze storingstermen in detail te bespreken, met name de oorzaken en de richtingen ervan en de waarschijnlijke omvang van de effecten ervan op de resultaten.

## **a) Difference-in-difference**

### *Motivering en identificatie*

Zoals gezegd, zal een eenvoudige vergelijking van begunstigden zelfs met een goed gekozen groep van niet-begunstigden waarschijnlijk geen valide evaluatie opleveren. Dit komt omdat het niet mogelijk is om het bestaan uit te sluiten van niet-geobserveerde verschillen tussen de beide groepen, die resulteren in een aanhoudend verschil in de uitkomsten, zelfs zonder de steun. Bovendien zal waarschijnlijk ook gewoonweg een vergelijking van de uitkomsten vóór



en ná de steun aan de begunstigden een onbruikbare evaluatie opleveren. Met deze methode kunnen de effecten van de steun niet worden afgezonderd van de effecten van andere factoren die ook een effect hebben op de uitkomsten voor de beide groepen, zoals de algemene economisch trend, veranderingen in het regelgevingskader of toenemende arbeidskosten.

Door de beide benaderingen echter te combineren kan het causale effect van de steun misschien wel worden gemeten: dit is de difference-in-difference-methode. Het achterliggende idee is het verschil in uitkomsten tussen de ondernemingen in de tijd te bezien. Reeds bestaande verschillen worden daarbij toegeschreven aan andere factoren dan de staatssteun. Alleen de veranderingen in deze verschillen (vandaar de term "difference-in-difference") worden dan aan de steun toegeschreven. Met andere woorden, deze methode vergelijkt het verschil in prestaties tussen de begunstigden van de steun en de controlegroep *vóór* de steunverlening en *ná* de steunverlening en schrijft dan dit verschil in veranderingen aan de steun toe. Deze methode functioneert indien zowel de begunstigden van de steun als de controlegroep in de tijd door andere factoren worden beïnvloed die de prestaties ook op dezelfde wijze beïnvloeden. Dan is de conclusie dat de steun de enige relevante factor is die de geobserveerde verandering in prestaties tussen de begunstigden en de controlegroep kan verklaren.

De cruciale aanname is dat de verschillen tussen de begunstigden van de steun en de controlegroep stabiel zijn in de tijd en dat beide groepen in die periode op dezelfde wijze door gemeenschappelijke schokken (afwijkingen van het gemiddelde) worden getroffen. Deze aanname kan in de praktijk wel eens niet gelden. Wanneer de begunstigden van de steun bijvoorbeeld kwetsbaardere ondernemingen zijn, is de kans groot dat zij sterker worden beïnvloed door economische terugval en het algemene ondernemingsklimaat. Daarom moet ook de controlegroep evenzeer kwetsbare ondernemingen bevatten. Al bij al is de keuze van de controlegroep van cruciaal belang voor de validiteit van deze methode. De identificatie van dit effect ligt dus niet bij het gebruik van differences-in-differences (hetgeen gewoonweg de technische toepassing is), maar in de juiste keuze van de controlegroep.

Bijzondere zorg is vereist bij het samenstellen van de controlegroep indien de ondernemingen die geen steun ontvangen, zelf hebben beslist om geen steun aan te vragen. Of al dan niet steun wordt aangevraagd, kan naar verwachting verband houden met het rendement dat het ontvangen van de steun oplevert. Daarom zijn er redenen om aan te nemen dat de verwachte uitkomsten van ondernemingen die geen steun aanvragen (in termen van banen, productiviteit, omzet enz.), verschillen van de voor de begunstigden verwachte uitkomsten. Indien bijvoorbeeld alle ondernemingen die steun aanvragen, bepaalde steun ontvangen, zijn de enige subsidiabele ondernemingen die geen aanvraag indienen, de ondernemingen die geen project hebben (in de aanname dat de kostprijs van een aanvraag laag ligt). Deze ondernemingen zullen dus waarschijnlijk niet alleen in absolute termen slechter presteren, maar naarmate de tijd verstrijkt ook in relatieve termen, terwijl de betere ondernemingen projecten uitvoeren en groeien. Banen, productiviteit of omzet kunnen niet verwacht worden parallel te blijven verlopen en dubbele differentiatie lost - meestal - het probleem niet op.

Daarom moeten ondernemingen in de controlegroep die geen steun genoten, zijn geselecteerd om redenen die geen invloed hebben op de gemeten uitkomsten. Zij mogen geen zelfselectie hebben gedaan en vrijwillig hebben besloten om niet deel te nemen. Het meest overtuigende opzet is wanneer niet deelnemen verband houdt met de niet-subsidiabiliteit die het gevolg is van een natuurlijk experiment. In dit geval is het weinig waarschijnlijk dat de niet-subsidiabiliteit toe te schrijven is aan niet-geobserveerde factoren die ook invloed hebben op de uitkomsten. Controlegroepen zouden bijvoorbeeld ondernemingen kunnen zijn in regio's die niet langer voor steun in aanmerking komen (indien de subsidiabiliteit geen verband houdt met hun eigen prestaties, maar eerder met een exogene gebeurtenis).

### *Toepassing*

Uit technisch oogpunt kunnen difference-in-difference-methoden worden toegepast hetzij binnen een model van lineaire regressie of bij matching. In het eerste geval wordt de controlegroep geselecteerd onafhankelijk van de observeerbare kenmerken en is deze dus al bij al vergelijkbaar met de volledige groep van de begunstigden van de steun. De observeerbare verschillen worden dan meegewogen bij een klassieke lineaire regressie. In het tweede geval bestaat de controlegroep uit ondernemingen die op basis van observeerbare factoren individueel vergelijkbaar zijn met elke gesteunde onderneming in de steekproef. De uitkomst voor iedere onderneming wordt vergeleken met de uitkomst van de onderneming(en) die het best met haar te vergelijken is (zijn), en de resultaten worden geaggregeerd. Deze beide methoden zijn twee verschillende methoden om de observeerbare verschillen te laten meewegen, maar er is geen fundamenteel verschil in termen van identificatie van het causale effect van het beleid.

Afhankelijk van de omstandigheden kan het de moeite lonen om een vergelijking te maken van de verschillen in uitkomsten van de begunstigden van de steun en de controlegroep vóór de steunverlening. Indien de uitkomsten al voordat de steun daadwerkelijk werd toegekend, systematisch beginnen af te wijken, bestaat de kans dat de controlegroep en de groep van de begunstigden al afweken om redenen die geen verband houden met de steun. In dat geval geeft de methode geen valide schatting van het causale effect van de steun. Dit is geen streng criterium voor de validiteit van de aanname: dit soort criterium bestaat niet. Toch is dit ten minste een nuttige, eerste eenvoudige test.

Andere methoden en robuustheidstests kunnen worden gebruikt wanneer er diverse potentiële controlegroepen bestaan die a priori valide zijn. De eerst en meest voor de hand liggende robuustheidstest is om verschillende difference-in-difference-schatters toe te passen en de resultaten te vergelijken. Daarnaast kan ook met deze verschillende controlegroepen worden gewerkt om zo tot een betrouwbaardere schatting te komen. Nemen we als voorbeeld een regeling voor het mkb in een bepaalde regio. Twee potentiële controlegroepen zijn de niet-mkb-bedrijven in die regio of het mkb in een aangrenzende regio. Geen van deze ondernemingen heeft er vrijwillig voor gekozen om geen steun aan te vragen: zij waren gewoonweg niet subsidiabel. Niettemin is geen van deze controlegroepen volmaakt: grote ondernemingen in dezelfde regio worden waarschijnlijk anders beïnvloed door algemene economische trends, terwijl het mkb in een aangrenzende regio misschien met andere

regionale schokken te maken krijgt. In plaats van te kiezen tussen deze twee mogelijke difference-in-difference-schatters kunnen zij ook worden gecombineerd en kan een triple difference (DDD) schatter worden toegepast: uitgaande van het "klassieke" difference-in-difference (DD) tussen mkb en niet-mkb in de betrokken regio, kan hetzelfde difference-in-difference in mindering worden gebracht bij de aangrenzende regio om de variatie in de uitkomsten van het mkb en niet-mkb in het steungebied uit te schakelen<sup>19</sup>. Een andere mogelijkheid zou zijn om systematisch te proberen een synthetische controlegroep samen te stellen, bestaande uit mkb-bedrijven uit verschillende aangrenzende regio's en niet-mkb-bedrijven uit dezelfde regio, om zo beter het patroon van de uitkomst voor de begunstigden van de steun vóór de steunverlening te repliceren (zie, voor nadere bijzonderheden, Abadie, Diamond en Hainmuller (2010)).

### *Inferentie*

Naast het belang van een zorgvuldige vormgeving en de keuze van een controlegroep, moet hier ook nader worden ingegaan op de kwestie van inferentie. Binnen deze context gaat het bij het begrip "inferentie" om de vraag of de effecten die werden geschat, werkelijk significant zijn. Statistische significantie is nog iets anders dan economische significantie. Dit laatste begrip verwijst naar de omvang van de geschatte effecten ten opzichte van de andere, uit oogpunt van economische theorie relevante parameters. Zoals hier reeds werd uiteengezet, is economische significantie van cruciaal belang. Toch is deze discussie in beginsel alleen relevant wanneer, uit statistisch oogpunt, de effecten voldoende precies worden geschat, d.w.z. dat valt uit te sluiten dat er helemaal geen effect is.

Er zijn geen redenen om aan te nemen dat een eenvoudige inferentie bij standaardaanname (zoals de homoscedastische aanname en de aanname dat er geen autocorrelatie is) waarschijnlijk een overschatting oplevert van de statistische significantie van de effecten.<sup>20</sup>

Het eerste probleem houdt verband met de clustering van data. Indien de controlegroep en de groep van de begunstigden van de steun telkens zeer homogeen zijn (ook al verschillen zij van elkaar), zullen alle ondernemingen binnen die groepen waarschijnlijk door dezelfde afwijkingen van het gemiddelde worden getroffen (schokken). Statistisch gesproken betekent dit dat de foutenterm een gemeenschappelijke component heeft. Indien de variantie van deze gemeenschappelijke component groot is ten opzichte van de variantie in voor individuele ondernemingen geobserveerde uitkomsten, zal de inferentie onzuiver zijn. Bij twee perioden en twee groepen kan dit probleem bijzonder ernstig zijn en een grensgeval vormen voor een identificatieprobleem: het is onmogelijk om het effect van de schokken die de verschillende groepen met elkaar gemeen hebben, af te zonderen van het beleidseffect. Dit probleem hoeft

---

19 Nemen we het voorbeeld van een regionale mkb-regeling voor het creëren van nieuwe banen. Stel dat bij afloop van de regeling blijkt dat in een bepaalde regio het mkb, in termen van banencreatie, 20% beter heeft gepresteerd dan grote ondernemingen. Indien in een vergelijkbare, aangrenzende regio (waar geen steun werd verleend) het mkb ook beter presteerde dan grote ondernemingen (met bijvoorbeeld 15%), dan kan het effect van de steun worden geschat op ruwweg 5%.

20 Dit punt is vooral beklemtoond in de context van de difference-in-difference-methode, maar dezelfde problemen kunnen zich ook voordoen bij de andere methoden die in dit werkdocument aan bod komen.

niet zo ernstig te zijn indien de groepen niet zo homogeen zijn. Wel moet steeds worden nagedacht over de aanwezigheid van gemeenschappelijke schokken voor homogene subsets van de groepen. Wanneer bijvoorbeeld de vraag lokaal is, moet normaal gesproken op het niveau van de gemeenten worden gecorrigeerd voor de clusterstructuur van de foutenterm. Hetzelfde kan ook gelden voor bedrijfstakken of sectoren.

Het tweede probleem dat zich voordoet, is wanneer paneldata worden gebruikt. Foutentermen bij de meeste data op ondernemingsniveau (zoals banen, productiviteit en investeringen) zijn normaal gesproken autogecorreleerd. Dit betekent dat afwijkingen van het gemiddelde in één periode, waarschijnlijk ook aanhouden in de volgende periode. Wordt dit punt verwaarloosd, dan kan dit leiden tot een overschatting van de nauwkeurigheid van de schatting van de effecten en tot het - vaker dan nodig - verwerpen van de hypothese dat het beleid geen effect heeft gehad. Dit kan een ernstig probleem zijn, zoals werd aangetoond in Bertrand, Duflo en Mullainathan (2004).

## **b) Instrumentele variabelen**

### *Motivering en identificatie*

Instrumentele variabelen zijn een klassieke methode die een antwoord biedt voor endogeniteit van verklarende variabelen. Omdat het ontvangen van steun kan worden gezien als een endogene verklarende variabele van de prestaties van een onderneming bij lineaire regressie, ligt het voor de hand om instrumentele variabelen te gebruiken voor de evaluatie van het effect van steun.

Een variabele is endogeen wanneer deze is gecorreleerd aan een niet-geobserveerd element, dat ook bepalend is voor de uitkomst. Nemen we bijvoorbeeld aan dat iemand probeert het effect van steunverlening op de werkgelegenheid bij ondernemingen te identificeren door regressieanalyse van de werkgelegenheid voor programmadeelname en andere observeerbare factoren. De aanname is nu dat het steunprogramma is gericht op zwak presterende ondernemingen die waarschijnlijk te maken hebben met moeilijke lokale marktomstandigheden. De marktomstandigheden zijn niet observeerbaar voor de evaluator en kunnen dus niet rechtstreeks worden gecontroleerd. Wanneer deze variabele echter buiten beeld blijft, zal het effect van de steun waarschijnlijk door de evaluator worden onderschat als gevolg van de endogeniteit van programmadeelname. Of de onderneming te maken heeft met gunstige of moeilijke marktomstandigheden, heeft ook een impact op zowel programmadeelname als op de werkgelegenheid, d.w.z. dat programmadeelname is gecorreleerd aan de foutenterm die werkgelegenheid verklaart. De impact van marktomstandigheden op programmadeelname betekent dat het niet mogelijk is om de correlatie tussen programmadeelname en werkgelegenheid volledig toe te schrijven aan het causale effect van de steun.

Er zijn echter ook andere factoren die wel de programmadeelname, maar niet de werkgelegenheid verklaren. Zoals bijvoorbeeld Criscuolo e.a. (2012) leerde, kan de geografische locatie bepalend zijn voor het totale bedrag dat in de regio voor het programma

beschikbaar is. Bovendien verandert de lijst van regio's die onder het programma valt, in de loop van de tijd. Indien het programmabudget voor een bepaalde regio in de loop van de tijd is veranderd om externe redenen (bijv. daling van het EU-gemiddelde voor het bbp per hoofd van de bevolking), dan heeft dit een effect op de programmadeelname, maar niet op de lokale marktomstandigheden van de onderneming. De verandering in werkgelegenheid die verband houdt met de exogene verandering in het toepassingsbereik van het programma, houdt geen verband met lokale marktomstandigheden. Door te focussen op dit "deel" van de variabele "programmadeelname" is het mogelijk om het reële effect van de deelname op de werkgelegenheid bij de onderneming te isoleren zonder interferentie van lokale marktomstandigheden. Dit is de achterliggende logica bij instrumentele variabelen.

Voor de evaluatie van staatssteun is een instrumentele variabele een variabele die kan verklaren dat de onderneming steun ontvangt, zonder dat zij een direct effect heeft op de andere niet-geobserveerde determinanten van de te meten uitkomst. Instrumentele variabelen bieden dan de mogelijkheid om scherp te stellen op deelname aan de regeling, zonder interferentie van de selectie-effecten. De logica achter instrumentele variabelen kan aan de hand van het volgende voorbeeld worden geïllustreerd.<sup>21</sup> In een eerste stap vindt voor programmadeelname een regressieanalyse plaats voor alle exogene variabelen, inclusief de instrumentele variabelen. In een volgende stap wordt de variabele "deelname" (die aangeeft of steun werd ontvangen) vervangen door de deelname zoals die in de eerste stap werd voorspeld: deze verwachte deelname is niet gecorreleerd aan het niet-geobserveerde element dat ook de uitkomst bepaalt.

#### *Problemen met zwakke instrumenten*

Een instrumentele variabele is een variabele die kan verklaren dat de onderneming steun ontvangt, zonder dat zij een direct effect heeft op de andere niet-geobserveerde determinanten van de te meten uitkomst. Deze eenvoudige en klassieke definitie laat echter een aantal praktische problemen buiten beeld. Er bestaan criteria om instrumenten op hun consistentie te testen wanneer meer instrumenten worden gebruikt dan strikt genomen noodzakelijk is om een model uit te werken. Toch bestaat er geen criterium voor de validiteit van instrumenten. In een studie die met instrumentele variabelen werkt, ligt de nadruk doorgaans vooral op het verklaren waarom van ieder afzonderlijk instrument mag worden aangenomen dat het niet gecorreleerd is aan de niet-geobserveerde determinanten van de prestaties van de onderneming (of het nu gaat om banen, productiviteit, omzet, investeringen enz.). Dit soort verklaringen die gebaseerd zijn op zowel economische argumenten als feitelijke gegevens, zijn noodzakelijk om de validiteit van de evaluatie te kunnen beoordelen. Toch zijn zij niet voldoende, met name wanneer met meerdere instrumenten wordt gewerkt.

Wanneer de kwaliteit van de instrumentele variabelen ter sprake komt, moet ook worden ingegaan op de problematiek van zwakke instrumenten, d.w.z. instrumenten die zwak

---

21 In de praktijk wordt de twee-staps kleinste-kwadraten (2SLS) schattingstechniek in één stap toegepast wegens welbekende inferentieredenen.

gecorrleerd zijn aan de uitkomstvariabele. Wanneer instrumentele variabelen zwak gecorrleerd zijn aan de endogene variabele, zijn de schattingen waarschijnlijk niet nauwkeurig. In die situatie kan de verleiding ontstaan om meer instrumentele variabelen toe te voegen. Het is een bekend gegeven dat door te werken met een voldoende groot aantal variabelen, het mogelijk is om de initiële variabele voldoende terug te halen om statistisch significante resultaten te verkrijgen. Tegelijk benadert de 2SLS-schattingstechniek steeds dichter de *biased ordinary least squares* (OLS) schattingsmethode.<sup>22</sup> Voor het risico op dit soort storing dient uitdrukkelijk een oplossing te worden gezocht bij iedere evaluatie die gebruikmaakt van de methode van de instrumentele variabelen. Met name moet een antwoord worden geboden voor de kwestie van de geloofwaardigheid, niet alleen van hun individuele, maar ook van hun gezamenlijke exogeniteit.

Een bijzonder geval is wanneer van de endogene variabele wordt aangenomen dat deze autogecorrleerd is. Indien van de bron van endogeniteit wordt aangenomen dat deze alleen gelijktijdig is, dan is het mogelijk om waarden uit het verleden als instrumentele variabelen te gebruiken. In dat geval zou wel moeten worden nagedacht over de precieze validiteit van deze benadering. Indien bijvoorbeeld verklarende variabelen autogecorrleerd zijn, zou dit ook het geval kunnen zijn voor de gemeten uitkomst. In dat geval zijn de vertraagde variabelen ook endogeen. Meer algemeen kunnen, indien de autocorrelatie van de verklarende variabelen zeer groot is, exogeniteitsaannames misschien niet gelden. Indien deze correlatie gering is, zou kunnen worden gebruikgemaakt van vele vertraagde variabelen (en potentieel ook toekomstige waarden) en zou men in de (hier reeds vermelde) valkuil lopen om te veel zwakke instrumenten te gaan gebruiken. Al bij al kan de methode om met waarden uit het verleden te werken, een valide strategie zijn maar zij moet wel met de nodige voorzichtigheid worden gehanteerd.

Over het algemeen gesproken is het, om de hierboven beschreven problemen te vermijden, sterk aanbevolen om slechts een beperkt aantal overtuigende instrumenten te gebruiken. Het is dan echter ook noodzakelijk om aan te tonen dat de instrumenten goede voorspellers zijn van de endogene verklarende variabele.<sup>23</sup>

#### *Variaties op twee-staps-schattingen: het selectiemodel van Heckman (1979)*

Wanneer de endogene variabele een treatmentvariabele is (deelnemingsdummy), kan de eerste regressie van de 2SLS gezien worden als een lineair probabiliteitsmodel van de kans op treatment. Dit lineaire probabiliteitsmodel is een lineaire benadering. In sommige gevallen

---

22 Een bijzonder boeiende praktische bespreking van de onzuiverheden die door zwakke instrumenten ontstaan, is te vinden in de bespreking door Bound, Jaeger en Baker (1995) van de statistische onzuiverheden in Angrist en Krueger (1991). Bovendien zijn schattingen met instrumentele variabelen onzuiver op eindige afstand. Daarom kunnen, zelfs met voldoende ruime datasets die kennelijke statistische significantie garanderen, niet-asymptotische onzuiverheden nog steeds aanzienlijk zijn.

23 Dit kan gebeuren door in de eerste stap van de regressieanalyse de Fisher-exact-toets te berekenen. Hoe hoger deze Fisher is, des te kleiner is de kans dat instrumenten zwak zijn. Stock, Wright en Yogo (2002) komen met een formeel criterium. Een instrument moet dan bijvoorbeeld in de Fischer exact-toets bij de eerste stap in de regressieanalyse groter zijn dan 10.

echter kan de kans op treatment, zelfs indien deze wordt beperkt tot subsidiabele ondernemingen, laag zijn. Daarom zijn lineaire benaderingen misschien te ruw om een reële benadering te kunnen zijn van deze kans op treatment en om te focussen op de staarten van de distributie, die net het punt zijn waar het precies om draait. Er zijn verschillende methoden om met deze problematiek om te gaan. Zij zijn alle erop gebaseerd om voor de kans op treatment het lineaire probabiliteitsmodel te vervangen door een niet-lineaire functie.<sup>24</sup>

Een klassieke benadering is om het evaluatieprobleem te behandelen in het kader van een selectiemodel (Heckman (1979)). In deze benadering wordt het selectie-effect behandeld als een probleem van een weggelaten variabele in de lineaire regressie van de uitkomst voor de observeerbare factoren en de deelname.<sup>25</sup> Op deze methode bestaan er diverse varianten, zoals bijvoorbeeld het hele model schatten aan de hand van maximale waarschijnlijkheid of de instrumentering van de steunverlening door de voorspelde waarde van de voor de selectie opgestelde vergelijking.<sup>26</sup>

Toch is het van cruciaal belang om goed na te denken over de identificatie en met name de keuze van variabelen. Het is niet voldoende om in beide stappen van de schatting dezelfde variabelen te gebruiken, zelfs al zijn de resultaten voldoende precies.<sup>27</sup>

Het is alleen betrouwbaar om een selectiemodel te schatten met een zogenaamde uitsluitingsvariabele. Een uitgesloten variabele is een variabele die de selectie van de begunstigde van de steun verklaart, maar niet de uitkomst. Het is niet voldoende om een variabele uit de basisvergelijking weg te laten om deze toe te voegen aan de lijst met verklarende variabelen voor de voor de selectie opgestelde vergelijking. Integendeel, deze

---

24 Dit onderdeel geeft een beknopte beschrijving van de selectiemodellen in dit onderdeel. Voor een completere presentatie zij hier verwezen naar de pionierstudie Heckman (1979) en vooral naar hoofdstuk 17 van Wooldridge (2002).

25 Deze weggelaten variabele is het verschil in de voorwaardelijke verwachting van de uitkomst voor de geselecteerde steekproef (hier dus de begunstigten van de steun). In bepaalde aannames over het selectieproces van de begunstigten van de steun (bijv. een probit- of logit-model) kan dit verschil formeel worden afgeleid (de omgekeerde Mills-ratio) en is deze een functie van de selectieparameters. In dat geval kan het effect van het beleid worden geïdentificeerd door de weggelaten variabele aan de regressie toe te voegen. De selectieparameters zijn onbekend, maar consistente parameters kunnen worden gevonden bij een eerstefase-schatting van het selectieproces. Dit leidt tot de schattingsprocedure die soms ook wel Heckit-methode wordt genoemd. In eerste instantie is daarvoor vereist dat de parameters worden gevonden die van belang zijn voor de selectie van de begunstigten van de steun, bijvoorbeeld een probit- of logit-schatting. Dan kan een consistente schatter van het beleidseffect worden verkregen door de geschatte omgekeerde Mills-ratio aan de lineaire regressie toe te voegen. Statistische softwarepakketten hebben meestal een functie om deze Heckman-schatting uit te voeren.

26 Voor de presentatie van al deze methoden, zij hier bijvoorbeeld verwezen naar hoofdstuk 17 van Wooldridge (2002).

27 Wanneer de voor de selectie opgestelde vergelijking niet-lineair is, is de omgekeerde Mills-ratio niet collineair met de andere verklarende variabelen, zelfs wanneer de eerste vergelijking alleen een subset van deze verklarende variabelen bevat. Dan is het model - in theorie - al gevonden. In dit geval geeft de omgekeerde Mills-ratio zeer vaak onvoldoende variatie te zien, hetgeen resulteert in zeer onnauwkeurige schattingen. Toch zou de schatting, zeker bij grote steekproeven, nog significante uitkomsten kunnen opleveren. Niettemin wordt, wanneer alle variabelen van het selectiemodel ook al in de basisvergelijking voorkomen, het model alleen geïdentificeerd door de niet-lineaire aspecten van een specifieke parametrische vorm.

uitsluitingsvariabele moet de selectie verklaren, maar mag geen effect hebben op de uitkomst waarvoor een verklaring wordt gezocht. Zij benadert in wezen erg dicht een valide instrumentele variabele. De keuze van dit soort variabele mag niet door gemakzucht zijn ingegeven, maar moet voortvloeien uit economische theorie, het institutionele kader en/of ervaring.

### c) Regression discontinuity design

Regression discontinuity design ("RDD") is de jongste uitbreiding van de evaluatietoolbox.<sup>28</sup> Het afgelopen decennium kende deze methode een groot succes in de academische wereld, vooral door haar eenvoud. In deze methode wordt gebruikgemaakt van het bestaan van een variabele die een discontinu effect heeft op de waarschijnlijkheid van beïnvloeding door een beleid. In het kader van staatssteunregels kunnen diverse soorten *discontinuity* nuttig zijn. In de eerste plaats zijn er geografische grenzen: de subsidiabiliteit in het kader van regelingen kan precies worden gekoppeld aan bestuurlijke grenzen, zoals gemeenten, NUTS-gebieden enz. In de tweede plaats zijn er de voorwaarden die worden opgelegd aan de ondernemingen die in aanmerking komen voor een regeling, met name wat betreft leeftijd en grootte.

Een voorbeeld: projecten die ondernemingen indienen, krijgen een puntenscore (op 100) en ondernemingen die ten minste 70 punten halen, krijgen steun, terwijl de overige ondernemingen geen steun krijgen. Een onderneming die 71 haalt, heeft een marginaal beter project dan een onderneming die 69 haalt. Het gevolg van dit marginale verschil is echter dramatisch: de ene onderneming krijgt steun en de andere helemaal geen steun. Een vergelijking maken van de uitkomsten voor deze beide ondernemingen, levert dus sterke aanwijzingen op voor het causale effect van de steun.

Formeel moet voor RDD de kans op het ontvangen van steun discontinu zijn, terwijl alle overige variabelen continu moeten zijn.<sup>29</sup> De technische uitvoering kan erg dicht in de buurt komen van de methode met instrumentele variabelen, waarbij een dummy voor drempeloverschrijding als instrument wordt gebruikt. Er zijn echter wel twee grote verschillen. In de eerste plaats is RDD op zwakkere aannames gebaseerd. Met name is niet a priori de onafhankelijkheid van het instrument vereist. Zo kunnen bijvoorbeeld bij een scoresysteem ondernemingen met een beter project meer aanvragen indienen dan ondernemingen met slechte projecten. De enige voorwaarde is dat de waarschijnlijkheid dat een aanvraag wordt ingediend, niet discontinu mag zijn rond de drempel. Het tweede verschil is dat de schattingen alleen berusten op ondernemingen die erg dicht bij de beide kanten van

---

28 Een formele en volledige beschrijving van RDD is te vinden bij Imbens en Lemieux (2008).

29 Formeel gesproken, zijn er twee verschillende soorten RDD: *sharp* RDD (SRD) en *fuzzy* RDD (FRDD). Bij *sharp design* (de methode die hier impliciet wordt beschreven) is er uitsluitend voor ondernemingen die een bepaalde drempel overschrijden, treatment. Bij *fuzzy design* is de discontinuïteit minder scherp: er is een discontinuïteit van de kans op treatment, maar deze verandert niet van 0 in 1. In absolute termen is, wat op subsidiabiliteitsvoorwaarden gebaseerde staatssteunregelingen betreft, het *design* alleen *sharp* indien de subsidiabiliteit wordt beschouwd als het treatment. Indien het ontvangen van steun echter het treatment is, gaat het om *fuzzy design*. Wanneer daarentegen de toekenning van steun op een scoresysteem is gebaseerd, wordt alleen rekening gehouden met onderneming die een aanvraag indienen, en is het *design* *sharp*.



de drempel blijven. Zwakkere aannames hebben dus consequenties: RDD-schattingen zijn dus nog lokaler dan schattingen met instrumentele variabelen doorgaans zijn. Indien de effecten van de steun verschillen voor ondernemingen die verder afstaan van de drempel, zijn RDD-schattingen geen correcte schatting van het effect op alle begunstigden van de steun.

De locatie van die schattingen kan van belang zijn wanneer effecten worden verwacht die sterk afwijken van de drempel. Bovendien zouden individuele ondernemingen aan de andere kant van de grens zeer sterk door het beleid kunnen worden beïnvloed. Dit zou bijvoorbeeld het geval kunnen zijn indien verplaatsingseffecten aanzienlijk zijn. In dat geval is het gebruik van RDD bij de geografische grens geen goede empirische strategie. Ten slotte ligt de kracht van RDD in de klemtoon op de enge bandbreedte rond de discontinuïteit. Indien de bandbreedte ruim is, mag van de impact van de overige kenmerken niet worden aangenomen dat deze constant is. Deze kwestie wordt normaal gesproken niet opgelost door te controleren op de observeerbare factoren, hetgeen een specifieke functionele vorm aanneemt.

Grafische inspectie van de data kan op dit punt de nodige garanties bieden wat betreft de betrouwbaarheid van de onderliggende aanname voor de RDD. Met name is het van bijzonder belang om op drie punten te controleren. In de eerste plaats moet er bij de drempel daadwerkelijk een discontinuïteit zijn in de steunverlening. Ten tweede moeten de te meten uitkomsten een discontinuïteit hebben op hetzelfde moment en geen andere gelijksoortige discontinuïteit elders. Ten derde moet ook worden nagegaan dat er geen discontinuïteit is bij de andere parameters die gecorreleerd zijn aan de uitkomst, met inbegrip van de *propensity* om steun aan te vragen.

Ten slotte kunnen misschien bewust discontinuïteiten worden gecreëerd om een evaluatie van de regeling mogelijk te maken. Met name zou het infaseren van beleid kunnen worden gebruikt om discontinuïteit te creëren en de effecten van een beleid te helpen identificeren.

### **C. Schattingen van structurele parameters**

In sommige gevallen is het mogelijk om een stap verder te zetten en om een theoretisch model (van bijvoorbeeld investeringen van ondernemingen) af te zetten tegen de data, om zo de sleutelparameters te vinden die van belang zijn. Deze benadering verschilt kwalitatief van de benaderingen die hiervoor werden gepresenteerd. Bij schattingen van structurele parameters wordt een volledig uitgewerkt theoretisch gedragsmodel van ondernemingen gehanteerd. Door schatting kunnen dan parameters worden gevonden die bepalend zijn voor het gedrag van de onderneming. Hiermee wordt de determinanten van het individuele gedrag van ondernemingen zo gericht mogelijk geëvalueerd en kunnen simulaties worden uitgevoerd voor de doelmatigheid van andere instrumenten. Toch vergt dit soort schattingen doorgaans meer middelen en data, maar stelt dit ook meer eisen in termen van aannames.

Het is niet mogelijk om precieze houvast te geven voor dit soort structurele schattingen omdat de identificatie, schatting en inferentie van geval tot geval moet worden afgeleid. Niettemin vinden de algemene aanwijzingen die in het voorgaande werden gegeven, ook hier toepassing. In de eerste plaats moet het theoretische model aansluiten bij de essentiële gestileerde feiten

van de markt. Daarnaast moet er ook een expliciet en correct antwoord zijn voor de kwesties van niet-geobserveerde kenmerken en van selectie.

#### **D. Verdere methodologische opmerkingen**

##### *Heterogeniteit van de treatmenteffecten*

In het vorige delen lag de nadruk op de schatting van het gemiddelde treatmenteffect op degenen onder treatment. De naam alleen al geeft aan dat er tussen de begunstigden verschillen in het effect van de steun zijn. Deze heterogeniteit kan vele oorzaken en vele gevolgen hebben. Het eerste gevolg kan zijn dat, indien steun erg doeltreffend is voor bepaalde ondernemingen maar veel minder voor andere, het gemiddelde effect statistisch misschien niet significant is. Dat een statistisch significant effect ontbreekt, betekent niet dat de steun geen effect heeft op ondernemingen. Uit beleidsoogpunt zijn de gemiddelde prestaties van een regeling een zeer interessante eerste indicatie. Toch is het, met het oog op de vormgeving van betere regelingen, even belangrijk om inzicht te proberen krijgen in de determinanten van deze heterogeniteit. Zo kan de aandacht worden toegespitst op ondernemingen waar de steun het doeltreffendst en het minst verstorend werkt.

Zo zou, telkens dat mogelijk is, het effect van de steun moeten worden geschat voor verschillende categorieën ondernemingen, zoals kleine/grote ondernemingen, jonge/oude ondernemingen, innovatieve ondernemingen, ondernemingen met een krap krediet enz.<sup>30</sup>

##### *Verstoren bij niet-gesteunde ondernemingen*

De (directe of indirecte) impact van de regeling op niet-deelnemers evalueren, is bijzonder verhelderend voor de evaluatie van staatssteun. Staatssteun kan markten verstoren door effecten op de niet-begunstigden ervan, bijvoorbeeld door kennis-spill-over van begunstigden of door verminderd relatief concurrentievermogen ten opzichte van begunstigden.

Bovendien kunnen de effecten op de niet-gesteunde ondernemingen of locaties een effect hebben op de validiteit van de evaluatie. Zo kan bijvoorbeeld een deel van het effect van regionale steun zich voordoen in de vorm van kansen langs de grens: ondernemingen die traditioneel aan de "verkeerde" kant van grens zijn gevestigd, verplaatsen hun vestiging gewoonweg naar de overkant. In dat geval zou een RDD bij de grens grotendeels dit verplaatsingseffect in beeld brengen en het risico inhouden op een overschatting van het reële geaggregeerde effect op het beleid. In dit soort situaties moet een andere empirische strategie worden gehanteerd (het kan bijvoorbeeld nuttig zijn om de robuustheid van de evaluatie voor ruimere regio's te controleren).

---

30 Een andere benadering zou erin kunnen bestaan om systematisch verschillende treatmenteffecten te schatten voor ondernemingen op verschillende punten van de conditionele distributie. In de literatuur is er een toenemend aantal studies die dit soort kwantiele treatmenteffecten schat, beginnend bij Abadie, Angrist en Imbens (2002). Dit is een bijzonder nuttig instrument om inzicht te krijgen in het intrinsieke karakter van de heterogeniteit van treatment. Uit strikt beleidsoogpunt is het echter minder bruikbaar, tenzij hiermee direct de verschillende ondernemingen kunnen worden bereikt afhankelijk van hun positie in de conditionele distributie.

## E. Data

Het is van cruciaal belang om toegang te hebben tot geschikte micro-economische data waarmee een effectbeoordeling kan worden uitgevoerd. Deze data moeten consistent zijn tussen begunstigden en niet-begunstigden. Daarom moeten zij van dezelfde bron afkomstig zijn, behalve natuurlijk de informatie over de steun zelf. De data moeten toegankelijk zijn op het meest verfijnde niveau, al kan in bepaalde gevallen in een later stadium een zekere vorm van aggregatie noodzakelijk zijn.

Er zijn data nodig die de resultaatindicatoren van zowel het treatment als de controlegroep in kaart brengen, met inbegrip van het tijdstip waarop de uitkomst wordt gemeten. Voorts zijn zoveel mogelijk data nodig over factoren die de uitkomsten en het besluit van de entiteiten om deel te nemen aan het steunprogramma, potentieel kunnen beïnvloeden. Deze data worden gebruikt om te controleren op verschillen tussen het treatment en de controlegroepen. Op ondernemingsniveau kan het bij deze data bijvoorbeeld gaan om locatie, grootte en demografische samenstelling, maar ook de bij de productie gebruikte input.

De meest voor de hand liggende databron is administratief, zoals fiscale balansgegevens of nationale enquêtes. Deze bronnen geven informatie over de locatie en activiteiten van ondernemingen, soms zelfs van individuele bedrijfslocaties. Hiermee kunnen meestal investeringen en omzet per activiteit in kaart worden gebracht en kunnen financiële kengetallen worden berekend. Ook grote nationale of EU-enquêtes zoals Community Innovation Surveys (CIS) kunnen van belang zijn. Zij bestrijken een grote en representatieve steekproef en geven zeer complementaire informatie over specifieke onderwerpen. Tot slot zijn ook geïntegreerde werkgever-werknemer-datasets een relevante informatiebron. Hiermee is het normaal gesproken mogelijk om arbeidskenmerken te koppelen aan iedere bedrijfslocatie. Dit kan cruciaal zijn wanneer de geografische dimensie van arbeid van belang is.

Naast indicatoren over resultaten en kenmerken van ontvangers van steun, zijn ook data over de steun en de procedure voor de toekenning van steun nodig. Deze informatie zal doorgaans van de steunverlenende autoriteit zelf komen. Hierbij gaat het om data over bedrag en tijdstip van toekenning van de steun aan de begunstigden. Toch zijn ook algemene data over de procedure voor de toewijzing van de steun nuttig. Data over afgewezen aanvragers zijn belangrijk, vooral wanneer de toekenning van de steun gebeurde op basis van een scoringsmechanisme.<sup>31</sup>

Toegang tot dit soort data is doorgaans gereguleerd. Daarom is het van essentieel belang ervoor te zorgen dat het wetenschappelijke team dat de evaluatie uitvoert, tijdig tot deze data toegang krijgt. Bovendien zijn deze administratie databronnen doorgaans met een zekere vertraging toegankelijk. Het is van belang om bij het uitwerken van het evaluatieplan rekening te houden met de beschikbaarheid van data.

---

31 De beschikbaarheid van data over afgewezen aanvragen is bijzonder waardevol voor studies waarin de regression discontinuity-benadering wordt gehanteerd.

## F. Voorbeelden

**Voorbeeld 1 (Regionale steun):** Criscuolo e.a. (2012)<sup>32</sup> hebben de Regional Selective Assistance-regeling ("RSA-regeling") in het Verenigd Koninkrijk geëvalueerd voor de periode 1986-2004. In deze periode kregen ondernemingen in achterstandsgebieden, in het kader van deze regeling, discretionaire subsidies. Deze regeling was de belangrijkste steunregeling voor bedrijven in het Verenigd Koninkrijk. De klijtlijnen voor steunverlening in het kader van de RSA-regeling waren getrokken door de richtsnoeren regionale steun, met name de kaarten met in aanmerking komende regio's ("regionale-steunkaarten"). Deze kaarten zijn over de jaren aangepast. Eenvoudig geformuleerd was het criterium om te bepalen of een regio voor steun in aanmerking kwam, de relatieve positie van de regio in termen van bbp per hoofd van de bevolking of werkloosheidspercentage. Daarmee kan de status van een regio veranderen omdat deze zich met de jaren heeft ontwikkeld of omdat het EU-gemiddelde voor het bbp per hoofd is gewijzigd (bijv. toen nieuwe lidstaten in 1995 tot de EU toetraden). Bovendien veranderen ook de voor het bepalen van de subsidiabiliteit gebruikte indicatoren met de jaren. Daarom is een deel van de verandering in de subsidiabiliteit van de ondernemingen niet afhankelijk van de situatie van de onderneming zelf, maar veeleer van gebeurtenissen die buiten het Verenigd Koninkrijk plaatsvonden of van veranderingen in de bestuurlijke regels. Door zich te concentreren op dit aspect van de veranderingen in subsidiabiliteit en na te gaan hoe deze veranderingen resulteerden in veranderingen in investeringsactiviteit, werkgelegenheid en productiviteit, hebben Criscuolo e.a. (2012) overtuigend de impact van de steun kunnen identificeren.<sup>33</sup>

**Voorbeeld 2 (Ondernemingssteun):** Martini en Bondonio (2012)<sup>34</sup> hebben twee gevallen van ondernemingssteun onderzocht: een investeringssubsidie die in heel Italië beschikbaar was (wet nr. 488/92) en diverse mkb-regelingen in de regio Piemonte. De eerste evaluatie is bijzonder interessant. Zij vergelijkt de ondernemingen die hun steunaanvraag gehonoreerd zagen (d.w.z. de begunstigden van de steun) met vergelijkbare ondernemingen die hun steunaanvraag afgewezen zagen omdat het voor steunverlening beschikbare budget zijn plafond had bereikt. Het is bijzonder nuttig om bij de evaluatie ook afgewezen aanvragen te gebruiken, zodat het selectie-effect niet gaat spelen dat zich meestal voordoet wanneer gewoonweg aanvragers en niet-aanvragers worden vergeleken. Deze groep ondernemingen had de eerste kwaliteitstoetsing doorstaan, hetgeen betekent dat zij een geloofwaardig investeringsproject hadden. Daarom deelden zij met de begunstigden van de steun dezelfde ambitie om te investeren in een geloofwaardig project. Niettemin hebben zij, door budgettaire beperkingen (rantsoenering), geen steun ontvangen. Het verschil in prestaties tussen (net wel)

32 Criscuolo, C., R. Martin, H.G. Overman en J. Van Reenen (2012), 'The causal effects of an industrial policy,' CEPR Discussion Papers 8818, C.E.P.R. Discussion Papers.

33 Technisch gezien gebruiken Criscuolo e.a. (2012) een benadering met instrumentele variabelen, zoals die hier in deze technische bijlage werd gepresenteerd.

34 Martini A. en D. Bondonio, 'Counterfactual impact evaluation of cohesion policy: impact and cost effectiveness of investment subsidies in Italy' Rapport voor DG REGIO (2012).

succesvolle aanvragers en (net wel) afgewezen aanvragers leverde een betrouwbare schatting op van het effect van steun.

**Voorbeeld 3 (Kredietgaranties):** Lelarge, Sraer en Thesmar (2010) evalueren de effecten van een regeling voor kredietgaranties in Frankrijk. Deze zogeheten Sofaris-regeling biedt kredietverstrekkers via garanties een verzekering tegen wanbetaling door kredietnemers. Kredietnemers betalen een verzekeringspremie, maar deze premie wordt gesubsidieerd. Lelarge, Sraer en Thesmar (2010) beschrijven uitdrukkelijk de aard van de selectie-effecten. In de eerste plaats is de kans groter dat ondernemingen met winstgevendere projecten bereid zijn de garantiepremie te betalen. In de tweede plaats zullen beheerders van de regeling waarschijnlijk maatschappelijk wenselijke projecten selecteren die anders misschien geen toegang tot particuliere financiering zouden krijgen. Al bij al doen ondernemingen aan zelfselectie voor het programma en vindt selectie ook plaats in de fase van de toekenning. Een en ander heeft waarschijnlijk invloed op de resultaten van naïeve evaluaties, op basis van bijvoorbeeld klassieke lineaire regressie of vergelijkingen met de best vergelijkbare onderneming.<sup>35</sup> Toch biedt de feitelijke en institutionele context van de regeling een bron voor het identificeren van de beleidseffecten. De regeling werd opgezet eind de jaren 1980 en was aanvankelijk beperkt tot ondernemingen die actief waren in de be- en verwerkende industrie en de zakelijke dienstverlening. In 1995 werden de overheidsmiddelen voor de regeling verhoogd en kwamen nieuwe bedrijfstakken in aanmerking (bouw, detail- en groothandel, vervoer, horeca en persoonlijke dienstverlening). Lelarge, Sraer en Thesmar (2010) vergelijken de nieuw subsidiabele ondernemingen met de voordien subsidiabele ondernemingen, om het effect van de regeling te meten op verschillende indicatoren zoals schulden, arbeidsplaatsen, kapitaalgroei, financiële onkosten en faillissementrisico. Ondernemingen in deze beide groepen zullen waarschijnlijk verschillen. Toch zouden ondernemingen vergelijkbare macro-economische schokken moeten voelen en zouden hun onderlinge verschillen dus waarschijnlijk niet in de tijd moeten veranderen, afgezien van de van het beleid zelf verwachte effecten.<sup>36</sup>

---

35 Dit is een voorbeeld waarin matching technieken - in dit geval één-op-één *nearest neighbour matching* - geen betere methode methode zijn om selectieproblemen op te lossen dan de kleinste kwadraten-schattingmethode (OLS). Zoals in deze technische bijlage reeds werd uiteengezet, bieden matching-technieken doorgaans geen oplossing voor de problematiek van selectie-effecten wanneer natuurlijke experimenten niet mogelijk zijn.

36 In de praktijk passen de auteurs een Heckman-selectiemodel toe met een uitsluitingsvariabele op ondernemingsniveau een klassieke strategie met instrumentele variabelen op sectorniveau. Voor meer details over deze methodieken, zie hoger in deze technische bijlage.

**Voorbeeld 4 (Creatief krediet):** Bakhshi e.a. (2013)<sup>37</sup> gebruiken een gerandomiseerd onderzoek met controlegroep (RCT) om het effect te meten van een steunregeling voor innovatieve bedrijven. De pilot, die in 2009 in Manchester van start ging, was zo opgebouwd dat vouchers (of "Creatieve kredieten") aselekt zouden worden toegekend aan mkb-bedrijven die een aanvraag indienden voor investeringen in creatieve projecten zoals de ontwikkeling van websites, de productie van video's en campagnes voor creatieve marketing, om te zien of zij een reëel effect op innovatie hadden. Creatieve kredieten deden écht nieuwe relaties ontstaan tussen het mkb en creatieve bedrijven, waarbij de toekenning van een creatief krediet de kans dat ondernemingen samen met een creatief bedrijf waarmee zij nog niet hadden samengewerkt, een innovatieproject zouden opzetten, met zeker 84% deed toenemen. Het onderzoek leerde dat de ondernemingen die creatieve kredieten kregen, op korte termijn een boost zagen in hun innovatie en omzetgroei in de zes maanden na de voltooiing van hun creatieve projecten. Deze positieve effecten bleven echter niet aanhouden en na twaalf maanden was er geen statistisch significant verschil meer tussen de groepen die de kredieten hadden gekregen, en de overige groepen. Volgens het rapport zouden deze resultaten niet in beeld zijn gekomen wanneer de normale door overheden gebruikte evaluatiemethoden waren toegepast. Het pleit ervoor om RCT's meer te gaan gebruiken bij de evaluatie van beleid dat de groei van bedrijven wil ondersteunen.

**Voorbeeld 5 (O&O&I-steun):** Einiö (2013) bestudeerde voor de periode 2000-2006 de impact van de O&O-subsidies van het Finse nationale innovatieagentschap Tekes op O&O-investeringen, werkgelegenheid en productiviteit. Tekes is belast met het merendeel van de O&O-steunverlening. De studie maakt gebruik van regionale variatie in het potentieel door Tekes aan O&O-steun toe te kennen budget dat resulteert uit hogere EFRO-financiering in delen van het noorden en oosten van Finland (doelstelling 1-gebieden). Deze gebieden waren aanvankelijk afgebakend tijdens de toetredingsonderhandelingen van Finland in 1995 en waren gebaseerd op een bevolkingsdichtheids criterium van maximaal 8 inwoners/km<sup>2</sup>. Door het relatief grotere budget voor O&O-steun was de kans om steun te ontvangen groter in doelstelling 1-gebieden dan in andere delen van het land. Dit veroorzaakte regionale variatie in treatment met een aanzienlijk hoger aandeel ondernemingen die werden gesteund in het doelstelling 1-gebied. Omdat de regionale toewijzing was gebaseerd op het bevolkingsdichtheids criterium dat vooraf was bepaald op basis van de bevolkingsdichtheid van 1993 (en niet op bijvoorbeeld verwachte toekomstige volumes O&O-investeringen of economische prestaties), controleert de studie voor de bevolkingsdichtheid van 1993 waarmee daadwerkelijk problemen in verband met regionale selectie worden geadresseerd. In de praktijk worden de treatmenteffecten geraamd met een instrumentele variabele-benadering waarbij een indicator voor doelstelling 1-regio wordt gebruikt als instrument voor toelating tot het programma. Deze benadering identificeert de impact van de steun op de ondernemingen

---

37 Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley en N. Rathbone, "Creative Credits. A randomised controlled industrial policy experiment", rapport voor Nesta, juni 2013, beschikbaar onder: [http://www.nesta.org.uk/sites/default/files/creative\\_credits.pdf](http://www.nesta.org.uk/sites/default/files/creative_credits.pdf).

die tot de steunregeling werden toegelaten als gevolg van hogere financiering in het doelstelling 1-gebied. De validiteit van de onderzoeksopzet wordt bevestigd door aan te tonen dat vóór het programma trends niet verschillen tussen ondernemingen die in het programma werden opgenomen, en de controlegroep. Einiö (2013) vindt positieve effecten op O&O-investeringen, werkgelegenheid en omzet bij de deelnemers die een O&O-subsidie kregen als gevolg van bijkomende geaggregeerde financiële O&O-steun in hun regio. Hoewel er geen onmiddellijke effecten op productiviteit zijn, levert de studie toch bewijs op voor productiviteitswinst op de lange termijn.

## Bijlage II: Lijst van mogelijke resultaatindicatoren

N.B. Dit is een indicatieve lijst die uitsluitend ter illustratie dient. De eigenlijke resultaatindicatoren dienen te worden bepaald in overeenstemming met de doelstelling van de steunregeling en die van de evaluatie.

### Directe impact van de steun op het niveau van de begunstigden

|  | RESULTAATDIMENSIE              | RESULTAATINDICATOREN  |
|--|--------------------------------|---|
| <b>Regionale steun</b>   | Positieve effecten             | Particuliere investeringen matchen overheidssteun<br><br>Toename werkgelegenheidsgroei in gesteunde ondernemingen   |
| <b>Steun voor onderzoek, ontwikkeling en innovatie (O&amp;O&amp;I)</b> | Bijkomende O&O&I-activiteiten  | Particuliere investeringen matchen overheidssteun<br><br>Bijkomende O&O&I-uitgaven door gesteunde ondernemingen<br><br>Aantal nieuwe onderzoekers in dienst bij gesteunde ondernemingen<br><br>Aantal geregistreerde nieuwe octrooien<br><br>Aantal ondernemingen dat wordt gesteund om nieuwe producten op de markten te brengen |
| <b>Milieu steun</b>  | Positieve milieu-effecten      | Verlaagde CO <sub>2</sub> -emissies van de begunstigde ondernemingen<br><br>Bijkomende productiecapaciteit hernieuwbare energie<br><br>Daling percentage gestort of verbrand afval<br><br>Aantal gesaneerde verontreinigde sites  |
|  | Vroege toepassing milieunormen | Percentage ondernemingen dat nieuwe milieunormen haalt ten minste X maanden/jaar voordat deze van kracht worden [aangezien ten minste 1 jaar wordt geëist en hogere steunintensiteiten zijn toegestaan indien dit sneller gebeurt dan 3 jaar]   |



|   |                                 |   |
|---|---------------------------------|---|
| <b>Steun voor energie(-<br/>infrastructuur)</b> | Verminderd energieverbruik      | Aantal huishoudens met verbeterde energieverbruikclassificatie<br><br>Daling jaarlijks primair energieverbruik openbare gebouwen<br><br>Aantal bijkomende energieverbruikers aangesloten op slimme netten |
|   | Steun voor hernieuwbare energie | Aandeel energie geproduceerd uit hernieuwbare energiebronnen  |
| <b>Risicofinanciering</b>                       | Positieve effecten              | Rendement behaald op het fonds<br><br>Hefboom particuliere investeringen<br><br>Aantal ondernemingen dat risicokapitaal ontvangt  |
|   | <i>Picking losers</i>           | Zwakke gemiddelde prestaties van ondernemingen waarin wordt geïnvesteerd door gebrekkig zakelijk management / onvoldoende participatie particuliere sector  |
|   | Onvoldoende diversificatie      | Te kleine / te regionaal beperkte fondsen met lage rendementsvooruitzichten die onaantrekkelijk blijven voor particuliere investeerders   |
| <b>Breedbandsteun</b>                           | Hogere breedbanddekking         | Hoger aandeel huishoudens met minstens 30 Mbps breedbandaansluiting<br><br>Hoger (gepland) aandeel huishoudens met minstens 100 Mbps breedbandaansluiting   |
|   | Efficiëntie                     | Investeringskosten/steun per aansluiting van een huishouden ( <i>homes passed</i> )<br><br>Aantal huishoudens dat zich op nieuwe diensten abonneert   |
| <b>Redding en herstructurering</b>              | Positieve effecten              | Behoud werkgelegenheid en activiteiten op ondernemingsniveau en op regionaal niveau<br><br>Veranderingen marktaandeel en productiviteit gesteunde ondernemingen   |
| <b>Luchtvaart</b>                               | Positieve effecten              | Aantal luchtvaartmaatschappijen dat de luchthaven gebruikt<br><br>Particuliere investeringen matchen overheidssteun<br><br>Toename regionale productiviteit en/of bruto toegevoegde waarde                |

|  |                    |   |
|--|--------------------|---|
|  | Negatieve effecten | <p>Duplicering verliesgevende infrastructuur of vliegroutes</p> <p>Verslechtering verkeersvolume bestaande infrastructuur (bijv. andere luchthavens in het verzorgingsgebied of andere vervoermiddelen)</p> |
|--|--------------------|---|

## Indirecte impact van de steunregeling

|   | RESULTAATDIMENSIE  | RESULTAATINDICATOREN  |
|---|--|---|
| Mogelijke positieve effecten                                  | Macro-economische voordelen  | Toename werkgelegenheid<br>Toename productiviteit en/of bruto toegevoegde waarde  |
|   | Diversificatie regionale economie  | Aantal sectoren volgens verschillende NACE-codes  |
|   | Toegenomen samenwerking particuliere en publieke sector                                  | Aantal ondernemingen dat samenwerkt met onderzoeksinstituten  |
|   | Positieve externaliteiten / overloopeffecten   | Aantal indirecte begunstigen (bijv. aantal derden met toegang tot de faciliteit)<br>Veranderingen werkgelegenheid of activiteit in andere bedrijven en regio's<br>(Luchtvaart) Aantal inwoners met betere vervoermogelijkheden in het verzorgingsgebied |
| Mogelijke negatieve effecten op mededinging en handelsverkeer | Sectorale bias   | In een multisectorale regeling steun in hoofdzaak toegekend aan één sector  |
|   | Bias tegenover verliesgevende ondernemingen of laagproductieve bedrijven (beletten exit) | Verhouding hoog-/laagproductieve ondernemingen  |
|   | Bias tegenover gevestigde ondernemingen  | Verhouding oude/jonge ondernemingen   |
|   | Versterken marktmacht  | Verandering marktmacht dominante speler   |
|   | Locatiegebonden effect   | Verhuizing van armere naar meer ontwikkelde regio   |
|   | Voor voorzieningszekerheid   | Lock-in in energiebronnen met hoog CO <sub>2</sub><br>Beoordeling of de zorg in termen van stroomuitval reëel zijn en blijven aanhouden<br>Afscherming nationale elektriciteitsmarkten  |
|   | Voor energie-infrastructuur  | Afscherming nationale elektriciteitsmarkten<br>Versterking marktmacht gevestigde onderneming  |

|  |                             |  |
|--|-----------------------------|--|
|  | Redding en herstructurering | Veranderingen werkgelegenheid of activiteit in andere bedrijven en regio's<br><br>Veranderingen marktaandeel en productiviteit gesteunde ondernemingen   |
|  | Luchtvaart                  | Duplicering verliesgevende infrastructuur of vliegroutes<br><br>Verslechtering verkeersvolume bestaande infrastructuur (bijv. andere luchthavens in het verzorgingsgebied of andere vervoermiddelen) |

### Bijlage III: Verklarende woordenlijst

|                           |   |
|---------------------------|---|
| <b>Controlegroep</b>      | Bij analyse aan de hand van een nulscenario komt het erop aan de meest vergelijkbare onderneming(en) of controlegroep te vinden, d.w.z. een groep ondernemingen die zo vergelijkbaar mogelijk is met de groep ondernemingen die de steun heeft ontvangen, afgezien van het feit dat zij die steun niét hebben gekregen.   |
| <b>Evaluatie</b>          | Het systematisch verzamelen en analyseren van informatie over programma's en projecten, de doelstelling ervan en de behaalde resultaten ( <i>delivery</i> ). Hieraan wordt kennis ontleend over de impact ervan, die als basis dient voor het doen van uitspraken. Evaluaties worden gebruikt om de doeltreffendheid te verbeteren en om besluiten over actuele en toekomstige programmatie te onderbouwen.   |
| <b>Impact</b>             | De verandering die geloofwaardig aan een interventie kan worden toegeschreven. Synoniem van "effect" van een interventie of "aandeel in de verandering".  |
| <b>Indicator</b>          | Een variabele die kwantitatieve of kwalitatieve informatie over een verschijnsel geeft. Meestal omvat deze een waarde en een meeteenheid.   |
| <b>Methode</b>            | Methoden zijn families van evaluatietechnieken en -instrumenten die voor verschillende doeleinden dienen. Meestal gaat het om procedures en protocollen die systematisering en consistentie borgen in de wijze waarop evaluaties plaatsvinden. Methoden kunnen focussen op het verzamelen of de analyse van informatie en data, kunnen kwantitatief of kwalitatief zijn, en kunnen proberen acties te beschrijven, te verklaren, te voorspellen of te onderbouwen. De keuze van methoden sluit aan bij de aard van de interventie, de evaluatievragen die worden gesteld en de wijze van onderzoeken - verklarend, verkennend, normatief enz. |
| <b>Nulmeting</b>          | De waarde van de indicator voordat de betrokken beleidsinterventie plaatsvindt.   |
| <b>Nulscenario</b>        | Om het effect van de steun op de begunstigde ondernemingen in te schatten, moet een nulscenario ( <i>counterfactual</i> ) worden geconstrueerd, d.w.z. een redelijk scenario moet worden uitgewerkt dat in beeld brengt wat waarschijnlijk met de ontvangers van de steun was gebeurd indien zij de steun niét hadden gekregen.   |
| <b>Resultaat</b>          | De specifieke dimensie van het welzijn van mensen waarop de beleidsactie inzet, d.w.z. waarvan wordt verwacht dat deze verandert dankzij de door een beleid ontworpen en uitgevoerde interventies. Voorbeelden zijn de mobiliteit in een gebied, de competenties in een bepaalde economische sector.  |
| <b>Resultaatindicator</b> | Een indicator die een specifiek aspect beschrijft van een resultaat, een meetbaar kenmerk. Voorbeelden zijn: de tijd om tegen een gemiddelde snelheid van A naar B te reizen, als aspect van mobiliteit; de resultaten van testen rond een bepaald thema, als aspect van competenties, of het percentage ondernemingen dat geen krediet tegen een bepaalde rente krijgt, als aspect van kredietrantsoenering door banken.   |

#### **Bijlage IV: Literatuurlijst**

Abadie, A., J. Angrist en G. W. Imbens (2002), 'Instrumental Variables Estimates of the Effect of Subsidised Training on the Quantiles of Trainee Earnings,' *Econometrica*, 70(1): 91-117

Abadie, A., A. Diamond en J. Hainmueller (2010), 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Programme,' *Journal of the American Statistical Association*, June 2010, Vol. 105, No 490, 493-505

Angrist, J. en A. Krueger (1991), 'Does Compulsory School Attendance Affect Schooling and Earnings,' *Quarterly Journal of Economics*, 106, 979-1014

Angrist, J. en Pischke (2008), "Mostly Harmless Econometrics: An Empiricist's Companion", Princeton University Press

Angrist, J. D. en Pischke (2010), 'The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics.' *Journal of Economic Perspectives*, 24(2): 3-30

Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley en N. Rathbone (2013), "Creative credits, a randomized controlled industrial policy experiment", Report for Nesta, beschikbaar onder: [http://www.nesta.org.uk/sites/default/files/creative\\_credits.pdf](http://www.nesta.org.uk/sites/default/files/creative_credits.pdf)

Bertrand, M., E. Duflo en S. Mullainathan (2004), 'How much should we trust differences-in-differences estimates?,' *The Quarterly Journal of Economics*, 119, 249-275

Bound, J., D. Jaeger en R. Baker (1995), 'Problems with Instrumental Variable Estimation When the Correlation Between the Instruments and the Endogenous is weak,' *Journal of the American Statistical Association*, 90(430): 443-450

Criscuolo C., R. Martin, H.G. Overman en J. Van Reenen (2012), "The causal effects of an industrial policy," CEPR Discussion Papers 8818, C.E.P.R. Discussion Papers

Duflo, E., R. Glennerster en M. Kremer (2007), "Using Randomisation in Development Economics Research: A Toolkit," CEPR Discussion Papers 6059, C.E.P.R. Discussion Papers

Duflo, E. en M. Kremer (2005), 'Use of Randomisation in the Evaluation of Development Effectiveness,' in *Evaluating Development Effectiveness*, ed. by O. Feinstein, G. K. Ingram, en G. K. Pitman. New Brunswick, New Jersey and London, U.K.: Transaction Publishers, vol. 7, p. 205{232}

Einiö, Elias (2013), 'R&D Subsidies and Company Performance: Evidence from Geographic Variation in Government Funding Based on the ERDF Population-Density Rule', *The Review of Economics and Statistics* (geaccepteerd voor publicatie)

Europese Commissie, *Evaluation Standards*, beschikbaar onder: [http://ec.europa.eu/dgs/secretariat\\_general/evaluation/docs/standards\\_c\\_2002\\_5267\\_final\\_en.pdf](http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/standards_c_2002_5267_final_en.pdf) [ook beschikbaar in het Frans en het Duits]

Garicano, L., C. Lelarge en J. Van Reenen, (2012), "Firm Size Distortions and the Productivity Distribution: Evidence from France," CEP Discussion Papers dp1128, Centre for Economic Performance, LSE

Givord, P. (2010), "Méthodes économétriques pour l'évaluation de politiques publiques", Document de travail de la DESE nr. G2010-08

Givord, P., Rathelot, R. en P. Sillard (2013), 'Place-based tax exemptions and displacement effects: An evaluation of the Zones Franches Urbaines programme', *Regional Science and Urban Economics*, 43(1): 151-163

Heckman, J. J. (1979), 'Sample Selection Bias as a Specification Error,' *Econometrica* 47, 153-161

Imbens, G. en J. Wooldridge (2009), 'Recent Developments in the Econometrics of Programme Evaluation,' *Journal of Economic Literature*, 47(1): 5-86

Imbens, G.W. en T. Lemieux (2008), 'Regression discontinuity designs: A guide to practice,' *Journal of Econometrics*, 142(2): 615-635

Lelarge, C., D. Sraer en D. Thesmar (2010), 'Entrepreneurship and Credit Constraints: Evidence from a French Loan Guarantee Programme,' NBER Chapters, in: *International Differences in Entrepreneurship*, p. 243-273, National Bureau of Economic Research, Inc.

Keane, M. P. (2010), 'A Structural Perspective on the Experimentalist School,' *Journal of Economic Perspectives*, 24(2): 47-58

Martini, A. en D. Bondonio (2012), "Counterfactual impact evaluation of cohesion policy: impact and cost effectiveness of investment subsidies in Italy", rapport voor Europese Commissie, DG Regio

Nederlandse Rijksoverheid (2012), "Durf te meten. Eindrapport Expertwerkgroep Effectmeting", beschikbaar onder: <http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2012/11/23/durf-te-meten-eindrapport-expertwerkgroep-effectmeting.html>

Nevo, A. EN M. D. Whinston, (2010), 'Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference?' *Journal of Economic Perspectives*, 24(2): 69-82.

Organisatie voor Economische Samenwerking en Ontwikkeling (OESO), *Evaluation Norms and Standards*, beschikbaar onder: <http://www.oecd.org/dac/evaluation/dcdndep/41612905.pdf>

Sims, C. A. (2010), 'But Economics Is Not an Experimental Science,' Journal of Economic Perspectives, 24(2): 59-68.

Stock, J., J. Wright and M. Yogo (2002): 'A Survey of Weak Instruments and Weak Identification in Generalised Method of Moments,' Journal of Business and Economic Statistics, 20(4): 518-29.

Verenigde Naties, United Nations Evaluation Group (2005), "Standards for Evaluation in the UN System", beschikbaar onder:  
[http://www.uneval.org/papersandpubs/documentdetail.jsp?doc\\_id=22](http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=22)

Wereldbank (2003), "Independent Evaluation: Principles, Guidelines and Good Practice", beschikbaar onder:  
<http://siteresources.worldbank.org/INTDGF/Resources/Evaluation&LearningNote.pdf>

Wooldridge, J. (2002), "Econometric Analysis of Cross Section and Panel Data", Cambridge (Mass.): MIT Press